Impact of Covid-19 on the human personality: An analysis based on document modeling

***Abstract***

*The outbreak of Covid-19 and the precautionary strategies have drastically impacted our lives in all dimensions. These may have an impact in our psychology too. In this study we have examined whether this pandemic have affected Five Factor Model personality traits using deep learning based document modeling. Using five different deep learning models we have analyzed tweeter data following the pretest posttest design.*

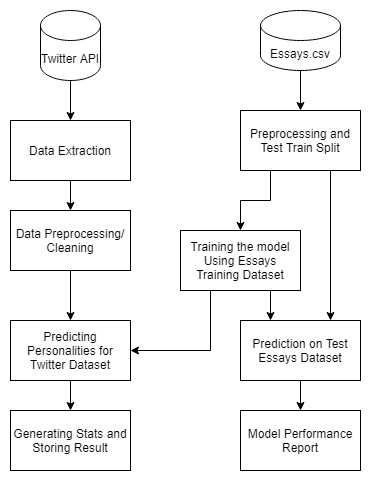
Introduction

Covid-19 has affected the globe terribly. The rapid spread of this virus and the precautionary measures to prevent it have impacted our lives in all dimensions. To control the spread most of the countries were put on lock down. To decelerate the spread such measures were essential. The anxieties over the virus along with the social restrictions have challenged the mental health and might have acute psychological consequences. Here, we have studied this matter.

Twitter is one of the most popular micro blogs on which people post real time messages about their opinions on a variety of topics. We have collected data from twitter to analyze the fact if this pandemic has affected people psychologically and whether there are any acute changes in Five Factor Model personality traits. We know that under normal circumstances, personality does not change abruptly over such a short period of time. But due to the extraordinary nature of this pandemic, and the drastic measures of control, however, personality may be reactive to these rapidly changing events. Because there are evidences from the psychopathology literatures [1, 2] indicating that traits change in response to distress and treatment for distress, respectively.

Personality is a complex combination of an individual’s motivations and psychological interactions with the environment. We have seen from the earlier literature that texts often reflect the author’s personality and using deep learning personality can be predicted with good accuracy [3, 5]. Moreover, casual texts like tweets, which are generally written in a natural way, might reflect individual’s personality more accurately. Here we have extracted personality traits from the tweets of the author. We have collected tweets of different types of users from different countries, ages, ethnicity and professions. Then we analyzed their tweets of last four months before this pandemic as well as four months after the pandemic individually. For training and testing we have used five different popular algorithms namely: Convolution Neural Network (CNN), Random Forest Classifier, Logistic Regression, Support Vector Machine with Radial Basis Function as kernel and Gaussian Naïve Bayes classifier. Five different networks for each of the five personality traits were trained. We have used the classical essays data set [4] to train and test the networks. As these models exhibited accuracy in line with the latest and state-or-the-art models presented in previous research [3, 5] so we have applied them to twitter data for our analysis.

Overview of our method



We have used Json for storing dictionaries and csv for storing datasets in different stages.

**Datasets Used**

Mainly we have used two datasets,

1. Stream-of-consciousness essays dataset:

This dataset is compiled by J.W. Pennebaker and Laura King [4]. It contains 2468 anonymous texts tagged with each of the five personality traits: EXT, NEU, AGR, CON, OPN. Each value is either ‘y’ or ‘n’.

1. Twitter Data extracted using Twitter API and Tweepy.

**Data Extraction:**

Due to the restricted access, we could collect the tweets of around 500 tweeter ids. Tweepy API was used for extracting the tweets. Then we found that most of the common peoples tweeted very rarely. We observed that celebrities, from different fields like music, films, politics, sports etc, tweet regularly. Thus we had to reject most of the ids and shorten the data set to 273 persons.

Then we compiled a dictionary for countries and their respective lockdown dates. We used this dictionary to classify the tweets into two categories:

1. **Pre\_Covid Tweets**: Tweets posted before COVID-19 Lockdown.
2. **Post\_Covid Tweets**: Tweets posted after COVID-19 Lockdown.

We maintained a gap of 15 days between the Pre\_Covid tweets and Post\_Covid tweets to avoid outlier cases.

**Data Cleaning and Preprocessing:**

1. Tweets contain URLs, Names/Mentions (e.g., @id\_begin\_referred) and Numbers. Since neither of these have a relation with emotion related to the person nor they contribute to the analysis, we have removed all of them using regular expressions and other preprocessing techniques.
2. Since in social media and micro blogs emojis play an important role to express the emotions, we converted the emojis to equivalent ASCII text. The emoji is generally used to express an emotion or feeling, but the ascii was literal representation of the emoji. Using a predefined emoji parser, we converted emoji to equivalent text. There we observed that text conversion was quite literal (e.g., sunglass\_face, face\_with\_sweat, heart\_eyes). So, we compiled a dictionary to convert the literal meaning of emoji to its equivalent emotion. We converted emoji associated with emotion to a text phrase. e.g. “sunglass\_face” to “cool”. We discarded emojis like Numbers, Flags, etc. Special Unicode characters were either converted into ASCII or they were removed.
3. Auxiliary verbs, stop words, Punctuations and extra spaces were removed. Also, words having length less than three were removed as almost all of them were punctuations or pronouns. Then using WordNet lemmatizer each token was lemmatized.
4. Then we removed those tweets which does not contain any emotion words. The list of emotion words is taken from NRC\_Emotion\_Lexicons

**Training:**

For training and testing, we have used stream-of-consciousness essays dataset compiled by J.W. Pennebaker and Laura King [4].

The text in essays is cleaned and preprocessed. Using the Keras Tokenizer, each word is assigned a number/token. Then all the texts were converted into sequence of numbers. Also, using pretrained Google’s Word2Vec, an embedding matrix for each word/token is created.

Models we used for learning.

1. Convolution Neural Network:

Architecture

{

Embedding Layer (pretrained weights).

Conv1D layer (filters = 128).

Conv1D layer (filters = 8).

Conv1D layer (filters = 256).

GlobalAvgPool1D Layer.

Dropout (rate = 0.5).

Dense (1, activation= sigmoid).

}

2) GaussianNB (priors= [0.5,0.5])

Prior Probability of all classes assuming to be equal.

3) LogisticRegression ()

4) RandomForestClassifier ()

5) SVM (kernel = ‘rbf’, gamma = 1.0, C = 10)

For CNN, we have used Tensorflow framework and rest of the models are applied using scikit-learn framework. For CNN we have used the pre\_trained\_embeddings directly in the embedding layer. For other models, we have taken mean of all word embeddings present in each text. For each of the traits we have created a separate model and trained and tested with its equivalent dataset.

**Validation:**

We have used ‘Leave One Out’ Cross Validation Technique to generate performance of each of the model.

Algorithm:

*For all i (0, N-1). Here N is 2468*

1. *Train a new model using all datapoints except ith datapoint.*
2. *Predict for ith datapoint using trained model.*
3. *Give Score for ith datapoint as 1 if prediction is correct else 0*

*Accuracy = Mean (Score )*

**Application:**

After training and testing, we used these five trained models on preprocessed twitter data. We analyzed the data for each twitter id and extracted their personality traits. For the same person, the status of the traits before the pandemic as well as after the pandemic was recorded. The change in values from Pre\_Covid to Post\_Covid for the same trait was taken into consideration. In this way, we calculated the change in personality traits for all traits and for each of the individuals.

Experimental Results

Each of the models was executed on the same data set to ensure the integrity of the result.

1. **Accuracy of the Models:** The accuracy of the models is shown in the Table 1. The given accuracy is based on the Leave One Out Cross Validation applied on the essays dataset. This accuracy is in line with the latest and state-or-the-art models presented in previous researches [3, 5].

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Traits | CNN | Gaussian NB | Logistic Regression | Random Forest Classifier | SVM\_RBF |
| 0 | cEXT | 94.94 | 59.26 | 58.32 | 61.67 | 67.62 |
| 1 | cNEU | 95.22 | 54.62 | 59.93 | 61.19 | 68.38 |
| 2 | cAGR | 96.44 | 57.76 | 58.12 | 62.70 | 67.66 |
| 3 | cCON | 95.45 | 57.82 | 58.59 | 58.63 | 66.58 |
| 4 | cOPN | 97.02 | 59.40 | 63.20 | 63.89 | 70.28 |

**Table 1: Training accuracy of the models**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Traits | CNN | Gaussian NB | Logistic Regression | Random Forest Classifier | SVM\_RBF |
| 0 | cEXT | 55.06 | 55.75 | 55.38 | 53.20 | 55.10 |
| 1 | cNEU | 55.59 | 53.64 | 56.92 | 54.82 | 56.76 |
| 2 | cAGR | 55.02 | 55.18 | 55.71 | 54.25 | 55.87 |
| 3 | cCON | 51.45 | 53.03 | 56.28 | 54.02 | 56.32 |
| 4 | cOPN | 58.34 | 58.63 | 61.70 | 59.27 | 61.70 |

**Table 1: Testing accuracy of the models**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| index | count | CNN | GaussianNB | Logistic Regression | Random Forest Classifier | SVM\_RBF |
| 0 | 0 | 49 | 161 | 108 | 75 | 69 |
| 1 | 1 | 86 | 85 | 103 | 89 | 95 |
| 2 | 2 | 86 | 20 | 39 | 89 | 74 |
| 3 | 3 | 38 | 5 | 19 | 18 | 29 |
| 4 | 4 | 14 | 2 | 4 | 2 | 6 |
| 5 | 5 | 0 | 0 | 0 | 0 | 0 |

1. **Number of persons who have exhibited change in their traits**: Number of persons who have exhibited change in their personality for each of the traits is shown in the Table 2. For each of the models we have tabulated the inclination of the change. Here, cEXT: pre\_1\_to\_post\_0 means with CNN we have found that in 45 persons there is a change in the trait extroversion from 1 to 0 (means extroversion to introversion), in 187 persons there is no impact on this particular trait and 41 persons have shown an inclination from introversion to extraversion.

**Table 2: Number of persons who have exhibited change in their traits**

1. **Number of traits changed:** Table 3 shows the count of persons having n number of traits changed in their personality n{0, 1, 2, 3, 4, 5} due to this pandemic. The interpretation follows like this: with CNN we have found that in 49 persons 0 numbers of traits got affected and in 86 persons 1 trait got affected.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Considering G,L,R,S** | **Considering C, L,R,S** | **Considering C, G ,R,S** | **Considering C, G,L,S** | **Considering C, G,L,R** |
| cEXT | cEXT | cEXT | cEXT | cEXT |
| pre\_1\_to\_post\_0 3 | pre\_1\_to\_post\_0 2 | pre\_1\_to\_post\_0 2 | pre\_1\_to\_post\_0 3 | pre\_1\_to\_post\_0 2 |
| no\_change 99 | no\_change 80 | no\_change 75 | no\_change 79 | no\_change 85 |
| pre\_0\_to\_post\_1 8 | pre\_0\_to\_post\_1 1 | pre\_0\_to\_post\_1 3 | pre\_0\_to\_post\_1 2 | pre\_0\_to\_post\_1 1 |
|  |  |  |  |  |
| cNEU | cNEU | cNEU | cNEU | cNEU |
| pre\_1\_to\_post\_0 0 | pre\_1\_to\_post\_0 0 | pre\_1\_to\_post\_0 0 | pre\_1\_to\_post\_0 0 | pre\_1\_to\_post\_0 0 |
| no\_change 152 | no\_change 106 | no\_change 105 | no\_change 125 | no\_change 126 |
| pre\_0\_to\_post\_1 0 | pre\_0\_to\_post\_1 0 | pre\_0\_to\_post\_1 0 | pre\_0\_to\_post\_1 0 | pre\_0\_to\_post\_1 0 |
|  |  |  |  |  |
| cAGR | cAGR | cAGR | cAGR | cAGR |
| pre\_1\_to\_post\_0 0 | pre\_1\_to\_post\_0 0 | pre\_1\_to\_post\_0 0 | pre\_1\_to\_post\_0 0 | pre\_1\_to\_post\_0 0 |
| no\_change 175 | no\_change 144 | no\_change 161 | no\_change 139 | no\_change 146 |
| pre\_0\_to\_post\_1 0 | pre\_0\_to\_post\_1 0 | pre\_0\_to\_post\_1 0 | pre\_0\_to\_post\_1 0 | pre\_0\_to\_post\_1 0 |
|  |  |  |  |  |
| cCON | cCON | cCON | cCON | cCON |
| pre\_1\_to\_post\_0 1 | pre\_1\_to\_post\_0 2 | pre\_1\_to\_post\_0 0 | pre\_1\_to\_post\_0 0 | pre\_1\_to\_post\_0 0 |
| no\_change 115 | no\_change 84 | no\_change 80 | no\_change 100 | no\_change 95 |
| pre\_0\_to\_post\_1 0 | pre\_0\_to\_post\_1 4 | pre\_0\_to\_post\_1 0 | pre\_0\_to\_post\_1 0 | pre\_0\_to\_post\_1 0 |
|  |  |  |  |  |
| cOPN | cOPN | cOPN | cOPN | cOPN |
| pre\_1\_to\_post\_0 0 | pre\_1\_to\_post\_0 1 | pre\_1\_to\_post\_0 0 | pre\_1\_to\_post\_0 0 | pre\_1\_to\_post\_0 0 |
| no\_change 110 | no\_change 102 | no\_change 102 | no\_change 112 | no\_change 110 |
| pre\_0\_to\_post\_1 2 | pre\_0\_to\_post\_1 0 | pre\_0\_to\_post\_1 0 | pre\_0\_to\_post\_1 0 | pre\_0\_to\_post\_1 0 |
|  |  |  |  |  |

**Table 3: Number of traits changed**

**4. Number of persons who exhibited change in trait:** Here in this table weare considering only those persons who have exhibited the same inclination of change in the particular trait considering at least four algorithms. Here considering G, L, R, S means Gaussian Naïve Bayes, Logistic Regression, Random Forest Classifier and SVM with RBF has shown the same result for the same person for that particular trait.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | CNN | GaussianNB | Logistic Regression | Random Forest Classifier | SVM\_RBF |
| 0 | cEXT : pre\_1\_to\_post\_0 | 45 | 28 | 26 | 40 | 36 |
| 1 | cEXT : no\_change | 187 | 200 | 201 | 187 | 181 |
| 2 | cEXT : pre\_0\_to\_post\_1 | 41 | 45 | 46 | 46 | 56 |
| 3 | cNEU : pre\_1\_to\_post\_0 | 53 | 14 | 10 | 30 | 36 |
| 4 | cNEU : no\_change | 175 | 248 | 254 | 219 | 205 |
| 5 | cNEU : pre\_0\_to\_post\_1 | 45 | 11 | 9 | 24 | 32 |
| 6 | cAGR : pre\_1\_to\_post\_0 | 31 | 8 | 30 | 7 | 8 |
| 7 | cAGR : no\_change | 200 | 258 | 221 | 251 | 239 |
| 8 | cAGR : pre\_0\_to\_post\_1 | 42 | 7 | 22 | 15 | 26 |
| 9 | cCON : pre\_1\_to\_post\_0 | 52 | 10 | 21 | 33 | 31 |
| 10 | cCON : no\_change | 172 | 255 | 224 | 192 | 198 |
| 11 | cCON : pre\_0\_to\_post\_1 | 49 | 8 | 28 | 48 | 44 |
| 12 | cOPN : pre\_1\_to\_post\_0 | 34 | 9 | 30 | 44 | 48 |
| 13 | cOPN : no\_change | 203 | 256 | 211 | 187 | 188 |
| 14 | cOPN : pre\_0\_to\_post\_1 | 36 | 8 | 32 | 42 | 37 |

**Table 4: Number of persons who exhibited change in traits**

Conclusion

We have applied five different algorithms to ensure the integrity of the results. Present research suggests mild personality change due to this pandemic. All five models have shown that in three persons we have observed a change in the trait extroversion. Two of them exhibited a change from extroversion to introversion whereas one has shown an inclination towards extroversion from introversion. For other traits there is no evidence of change. While considering four models out of five we observed almost similar results (see Table - 4) extroversion being the most affected trait. Before this experiment, we had a premonition that Neuroticism might be the most affected trait. Because we have seen in the earlier literature [1] that when individuals go through great amount of distress or depressive episode, Neuroticism tends to increase. But in contrast to our belief, we found that Neuroticism did not increase due to this pandemic. In case of the other traits namely Openness, Agreeableness and Conscientiousness there was no evidence of significant change considering all possible combination of four models.

References

**[1]** Karsten J, Penninx BW, Riese H, Ormel J, Nolen WA, Hartman CA. *The state effect of depressive and anxiety disorders on big five personality traits*. JPsychiatr Res. 2012;46:644–50. pmid:22349302

**[2]** Roberts BW, Luo J, Briley DA, Chow PI, Su R, Hill PL. *A systematic review of personality trait change through intervention*. Psychol Bull. 2017;143:117–41. pmid:28054797

**[3]** Navonil Majumder, SoujanyaPoria, Alexander Gelbukh, Erik Cambria. “*Deep Learning-Based  
Document Modeling for Personality Detection from Text”,* IEEE Intelligent Systems (Volume: 32, Issue: 2, Mar.-Apr. 2017)

**[4]** J.W. Pennebaker and L.A. King, “*Linguistic Styles: Language, Use as an Individual Difference*,” J. Personality and Social Psychology*,* vol. 77, no. 6, 1999, pp. 1296–1312.

**[5]** Nazar Akrami, Johan Fernquist, Tim Isbister, Lisa Kaati, Björn Pelzer, “*Automatic Extraction of Personality from Text: Challenges and Opportunities*”, 2019 IEEE International Conference on Big Data (Big Data).

**[6]**Angelina R. Sutin ,Martina Luchetti,Damaris Aschwanden,Ji Hyun Lee,Amanda A. Sesker,Jason E. Strickhouser,Yannick Stephan,Antonio Terracciano. *Change in five-factor model personality traits during the acute phase of the coronavirus pandemic*