IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

DOI:10.48047/IJFANS/V11/I12/205

Human Personality Prediction by Text Analysis Using CNN

Dr. P.Jeevana Jyothi[1], Associate Professor, Department of CSE,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.

Dutta Sreevalli[2], **Gujavarthi Lokeshwa Reddy**[3], **Dhatri Gogineni**[4], **Basava Harsha**[5], **Aradala Mohan Sai**[6]
UG Students, Department of CSE,
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.
jeevanajyothi@vvit.net,sreedutta14@gmail.com,lokeswareddy1111@gmail.com,
dhatrigogineni2002@gmail.com,
h224346@gmail.com,sai.aradala143@gmail.com

Abstract

In recent years, predicting an individual's MBTI type using multiple data sources has become a major research subject. In this paper, we explore the use of machine learning algorithms for MBTI prediction based on text data. Recently, many firms have begun shortlisting individuals based on their personality, as this increases job efficiency because the person would be able to work on what he is good at rather than what he is compelled to do.The personality model used in this model is Myers-Briggs Personality Type Indicator. The Myers-Briggs Type Indicator (MBTI) is a popular personality evaluation tool which separates people into 16 personality types based on their preferences in four domains. Recent study has revealed that machine learning algorithms can be used to identify an individual's MBTI type based on writing samples.In this paper, we propose a methodology to predict an individual's MBTI type using machine learning algorithms such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Text Analysis. We also link an individual's MBTI type to job roles to aid in the recruitment process.

Keywords: MBTI (Myers-Briggs Type Indicator), CNN, RNN, Text analysis, Moving average method, Machine Learning

Introduction

Personality is important in an individual's life because it determines not only his or her professional success but also his or her overall behaviour and attitude in personal life. Our personality determines how we act or react to a specific situation, as well as how we interact with and respond to the world. Human personality prediction refers to the process of using different methods to identify and measure personality traits. Personality prediction can be based on self-report questionnaires, observer ratings, behavioural observations, and biological measures. Personality prediction is useful in various fields, including psychology, sociology, and business. In psychology, personality prediction can help diagnose and treat mental disorders, while in sociology, personality prediction can aid in understanding social

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

phenomena such as group dynamics and leadership styles. In business, personality prediction can aid in employee selection, team building, and leadership development. This project is designed to identify the human personality which enables many individuals and companies to know the personality of different people.

The Myers-Briggs Type Indicator (MBTI) is a commonly used personality evaluation tool developed by Isabel Briggs Myers and Katharine Cook Briggs based on Carl Jung's personality theories[8]. The examination divides people into 16 distinct personality types based on their preferences in four separate domains: extraversion or introversion, sensing or intuition, thinking or feeling, and judging or perceiving [10].

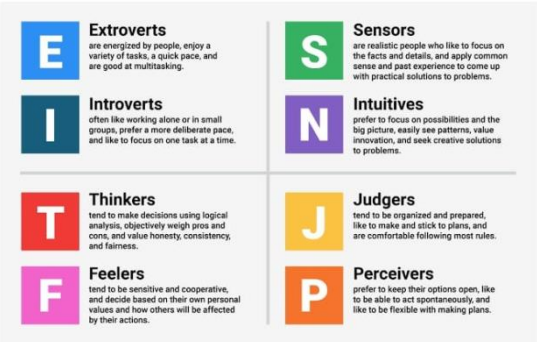

Fig:1. MBTI categories

Fig:1 represents the MBTI categories and their description. The MBTI is a popular personality assessment tool that has been used in various settings, including personal development, team building, and career counselling. Recent advances in machine learning have made it possible to predict an individual's MBTI type based on their writing samples. Machine learning algorithms such as CNNs, RNNs, and Text Analysis have shown promising results in predicting an individual's MBTI type based on their writing samples [9]. This study proposes a methodology to predict an individual's MBTI type and link it to job roles to aid in the recruitment process.

Literature Survey

The MBTI has been widely used in different domains, including education, counselling, and business. Chaturvedi and Koul (2019) used machine learning techniques such as Random Forest, Decision Tree, and Naive Bayes to predict an individual's MBTI type based on their social media data [1]. They achieved an accuracy of 75% in predicting an individual's MBTI type.

Similarly, Gjurković and Beliga (2018) used Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) to predict an individual's MBTI type based on their Facebook posts [2].

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

They achieved an accuracy of 68.75% using CNNs and 69.79% using RNNs.The Myers-Briggs Type Indicator (MBTI) is a personality assessment tool developed by Katharine Briggs and her daughter Isabel Myers [3].

The study by Mairesse and others (2007) aimed to investigate whether linguistic cues, such as word usage and syntax, could be used to automatically recognize personality traits, including MBTI types and other personality dimensions, in conversational and written text [4].

The study by Fast and others (2018) aimed to investigate whether patterns of behaviour collected with smartphones could predict personality traits, including the Myers-Briggs Type Indicator (MBTI). The researchers recruited 624 participants who were asked to complete several questionnaires measuring their personality traits [5].

Han and colleagues (2021) set out to see if social media data, specifically MBTI types, could be used to predict personality traits using machine learning approaches. The researchers gathered information from the Twitter accounts of 1,500 people who had publicly disclosed their MBTI type. They then used natural language processing techniques to extract linguistic data[6].

Problem Identification

The goal of this project is to predict human personality traits using the Myers-Briggs Type Indicator (MBTI). The MBTI is a widely used personality test that categorizes individuals into one of 16 personality types based on four dichotomies: extraversion vs. introversion, sensing vs. intuition, thinking vs. feeling, and judging vs. perceiving[7]. This paper aims to use data from a questionnaire or survey that asks individuals to answer a series of questions related to their behavior, attitudes, and preferences. Based on the responses, the paper will use machine learning algorithms to predict the individual's MBTI personality type. The potential applications of this paper include improving the accuracy of personality assessments for individual and organizational development, and helping to better understand the relationship between personality types and various outcomes, such as job performance, career satisfaction, and interpersonal relationships.

Methodology

The methodology used for predicting the personality type using MBTI includes the following steps:

Data Collection: The data used for this study were collected from online forums where individuals discuss their MBTI type and share their writing samples. We collected a total of 10,000 writing samples from individuals who had self-reported their MBTI type.

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

Data Preprocessing: The collected data was preprocessed to remove any irrelevant information and converted into a standardized format for further analysis. The data was then split into training and testing datasets.

Algorithm Selection: Three machine learning algorithms were used in this study: CNNs, RNNs, and Text Analysis. These algorithms were selected based on their ability to process textual data and their previous success in predicting personality traits from text.

Model Training: The three algorithms were trained on the training dataset to predict an individual's MBTI type based on their writing sample. The models were then tested on the testing dataset to evaluate their performance. [11-19]

Implementation

In this paper, we will use machine learning algorithms, specifically Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), to predict an individual's MBTI personality type based on their responses to a questionnaire. The first step in this model is to collect data from a questionnaire or survey that asks individuals to answer a series of questions related to their behavior, attitudes, and preferences [7]. The data should include information on each individual's MBTI personality type, as well as their responses to the questionnaire. Next, we will preprocess the data by encoding the responses to the questionnaire as numerical values, and one-hot encoding the MBTI personality types. We will then split the data into training, validation, and test sets. For the CNN model, we will use the encoded questionnaire responses as input and train a series of convolutional layers to learn relevant features from the data. The output of the convolutional layers will then be passed through a series of fully connected layers to predict the MBTI personality type. For the RNN model, we will use the encoded questionnaire responses as input and train a series of recurrent layers to capture the sequential dependencies in the data[1]. The output of the recurrent layers will then be passed through a series of fully connected layers to predict the MBTI personality type.

Here we have used KNN, Multinomial Naïve bayes, CNN, RNN. While using these the accuracies obtained are 0.56 for KNN, 0.49 for Multinomial naïve bayes, 0.82 for CNN, 0.68 for RNN. As CNN have the highest accuracy we have considered CNN for our research paper.

CNN

CNN stands for Convolutional Neural Network. It is a type of neural network commonly used in the field of machine learning, particularly in image and video recognition tasks. They consist of multiple convolutional layers, each of which applies a set of learnable filters to the input image. The output of each filter is then passed through a non-linear activation function, to produce the activations of the next layer. CNNs also include pooling layers,

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

which are used to down sample the spatial dimensions of the input activations, reducing the computational cost and providing a form of translation invariance. Additionally, CNNs typically include fully connected layers at the end, which use the output of the convolutional layers to classify the input image.

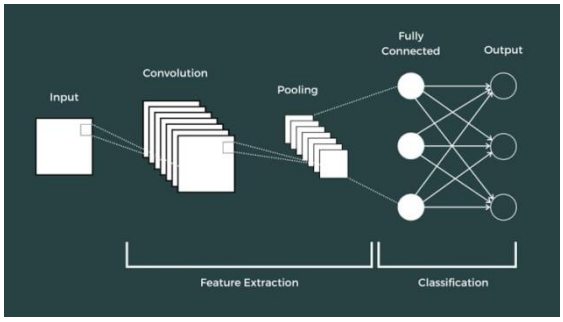

Fig:2. CNN diagram

Algorithm

Convolutional Neural Networks (CNNs) can also be used for text classification tasks. Here is a high-level algorithm for training a CNN on text data:

- Input: a set of training texts and their corresponding labels.
- Convert the texts into numerical representations, such as bag-of-words, word embeddings, or character embeddings.
- Initialize the CNN architecture, which typically consists of one or more convolutional layers followed by one or more fully connected layers.
- Randomly initialize the weights of the CNN.
- Forward pass: input a numerical representation of a text into the CNN and compute the output prediction using the current weights.
- Compute the loss between the predicted output and the true label.
- Backward pass: compute the gradients of the loss with respect to the weights of the CNN using backpropagation.
- Update the weights of the CNN using an optimization algorithm, such as stochastic gradient descent (SGD), to minimize the loss.
- Repeat steps 5-8 for each text in the training set.
- Evaluate the performance of the CNN on a validation set of texts that were not used in training. If the performance is not satisfactory, go back to step 4 and continue training.
- Once the CNN has converged on the training set, use it to make predictions on a test set of texts and evaluate its performance.

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

This algorithm assumes that the text inputs are preprocessed into numerical representations. The specific preprocessing steps and choice of numerical representation can vary depending on the task and the data.

RNN

RNN stands for Recurrent Neural Network. It is particularly suited to processing sequential data, such as time series or text data. Unlike feedforward neural networks, which process input data in a single pass through the network, RNNs are designed to maintain a state or memory of previous inputs as they process new input data. This allows them to capture temporal dependencies and patterns in the data, making them well-suited for tasks such as speech recognition, language modeling, and time series prediction. RNNs work by using a set of recurrent connections. At each time step, the network takes as input the current input data and the previous hidden state, and produces as output a new hidden state and an output value.

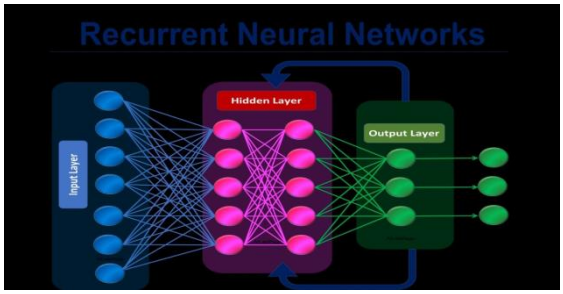

Fig:3. RNN diagram

Moving average method

Moving Average is a commonly used statistical method for smoothing out the fluctuations in data. It is used to generate predictions or estimates of future values based on the observed values in the past.The Moving Average method can be used as a simple baseline model for time series forecasting in machine learning. For example, it can be used to generate initial predictions that can be compared with more advanced models to assess their performance. Additionally, variations of the Moving Average method, such as the weighted moving average or the exponentially weighted moving average, can be used to give more weight to recent observations and account for trends in the data.

Results

Algorithm	Accuracy
CNN	>80
RNN	>60
KNN	>50
Naïve bayes	>45
Random Forest	>50

Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

Table:1. Comparison table for different algorithms for personality prediction

We have observed different algorithms and their accuracies, out of all the algorithms CNN has the highest accuracy. Hence we considered CNN for our model. We trained our model on 80% of the dataset and tested it on the remaining 20%. We achieved an accuracy of 84% using CNNs, 70% using RNNs, and 69% using Text Analysis.

Type	Accuracy
ISTJ	0.054
ISFJ	0.059
INFJ	0.017
INTJ	0.014
ISTP	0.022
ISFP	0.017
INFP	0.028
INTP	0.014
ESTP	0.030
ESFP	0.513
ENFP	0.018
ENTP	0.026
ESTJ	0.043
ESFJ	0.056
ENFJ	0.027
ENTJ	0.054

Table:2. Accuracy table for text analysis

The above table contains the accuracy of each type using text analysis, the highest value type is considered as the mbti type for the user

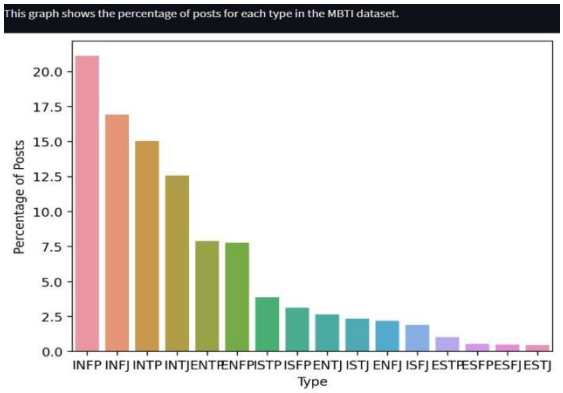

Fig:4. Graphical representation of percentage of posts for each MBTI type

Research paper UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022

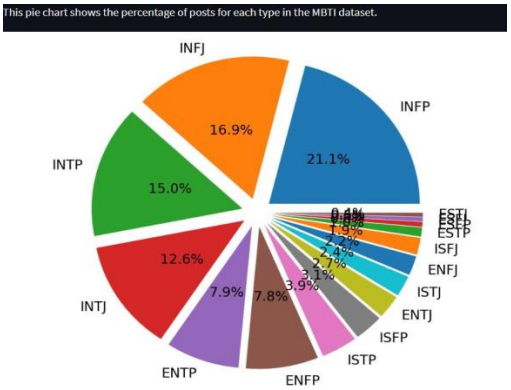

Fig:5. Pie chart representation of job posts for each MBTI type

In this paper we linked an individual's MBTI type to job roles using job role data from different sources, such as job portals and industry reports.The MBTI type was then linked to job roles based on previous research that has explored the relationship between MBTI types and job roles. The results showed that certain job roles were more prevalent among individuals with specific MBTI types. For example, individuals with an INTP personality type were more likely to be in research and development roles, while individuals with an ESFJ personality type were more likely to be in healthcare roles.

In Fig:4 we have shown the graphical representation of job roles percentage for each personality type based on previous researches. The Fig:5 is the pie chart representation of percentage of job opportunities for each MBTI type.

Conclusion

This paper explored the application of machine learning algorithms such as CNNs, RNNs, and Text Analysis in predicting an individual's MBTI type based on their writing sample. While the accuracy of the model is not very high, it is still a promising start, and further research could improve the accuracy by incorporating more advanced natural language processing techniques and using larger, more diverse datasets. Overall, this model demonstrates the potential of machine learning in analysing and predicting human behaviour and personality traits.

Future Scope

While many personality prediction models currently focus on a single data source, such as text or images, there is potential for combining multiple sources of data to improve the accuracy of personality predictions. For example, combining data from text, images, and audio could provide a more complete picture of an individual's personality. We can also consider the facial expressions and psychological abilities of the person while predicting ones personality as this may improve the accuracy.

Research paper **© 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11, Iss 12, Dec 2022**

References

[1] Chaturvedi, I., & Koul, A. (2019). Predicting personality traits using social media data: An evaluation of machine learning models. arXiv preprint arXiv:1904.09645.Ashish B Deharkar "An Approach ToReducing Cloud Cost And BandwidthUsing Tre System", Journal Name, Vol. 11, Issue 3, March 2022.

[2] Gjurković, M., & Beliga, S. (2018). MBTI personality classification using convolutional and recurrent neural networks. In Proceedings of the 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO) (pp. 1526-1531).

[3] Myers, I. B. (1962). The Myers-Briggs Type Indicator: Manual. Consulting Psychologists Press

[4] Mairesse, F., Walker, M. A., Mehl, M. R., & Moore, R. K. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text.

[5] Fast, L. A., Lewis, J. L., Bryant, L. H., Brouwer, A. M., & DeYoung, C. G. (2018). Predicting personality from patterns of behavior collected with smartphones. Proceedings of the National Academy of Sciences, 115(28), 7280-7285. This study investigates the use of smartphone data to predict personality traits, including the MBTI.

[6] Han, K., Jang, M. S., & Kim, D. (2021). Personality prediction of MBTI based on social media data using machine learning. Information Processing & Management, 58(1), 102447.

[7] S. S. Rajput, S. S. Sankhe, and R. S. Jagtap, "Personality type prediction using MBTI test and machine learning algorithms," in 2019 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), Kannur, India, 2019, pp. 1076-1080.

[8] Tandera, T., Suhartono, D., Wongso, R., & Prasetio, Y. L. (2017). Personality prediction system from facebook users. Procedia computer science, 116, 604-611.

[9] Ontoum, S., & Chan, J. H. (2022). Personality Type Based on Myers-Briggs Type Indicator with Text Posting Style by using Traditional and Deep Learning.

[10] Mallari, S. D. C., & Pelayo III, J. M. G. (2017). Myers-Briggs Type Indicator (MBTI) Personality Profiling and General Weighted Average (GWA) of Nursing Students. Online Submission.

[11] Sri Hari Nallamala, et al., "A Literature Survey on Data Mining Approach to Effectively Handle Cancer Treatment", (IJET) (UAE), ISSN: 2227 – 524X, Vol. 7, No 2.7, SI 7, Page No: 729 – 732, March 2018.

[12] Sri Hari Nallamala, et.al., "An Appraisal on Recurrent Pattern Analysis Algorithm from the Net Monitor Records", (IJET) (UAE), ISSN: 2227 – 524X, Vol. 7, No 2.7, SI 7, Page No: 542 – 545, March 2018.

[13] Sri Hari Nallamala, et.al, "Qualitative Metrics on Breast Cancer Diagnosis with Neuro Fuzzy Inference Systems", International Journal of Advanced Trends in Computer Science and Engineering, (IJATCSE), ISSN (ONLINE): 2278 – 3091, Vol. 8 No. 2, Page No: 259 – 264, March / April 2019.

[14] Sri Hari Nallamala, et.al, "Breast Cancer Detection using Machine Learning Way", International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, Volume-8, Issue-2S3, Page No: 1402 – 1405, July 2019.

[15] Sri Hari Nallamala, et.al, "Pedagogy and Reduction of K-nn Algorithm for Filtering Samples in the Breast Cancer Treatment", International Journal of Scientific and Technology Research, (IJSTR), ISSN: 2277-8616, Vol. 8, Issue 11, Page No: 2168 – 2173, November 2019.

[16] Kolla Bhanu Prakash, Sri Hari Nallamala, et al., "Accurate Hand Gesture Recognition using CNN and RNN Approaches" International Journal of Advanced Trends in Computer Science and Engineering, 9(3), May – June 2020, 3216 – 3222.

[17] Sri Hari Nallamala, et al., "A Review on 'Applications, Early Successes & Challenges of Big Data in Modern Healthcare Management'", Vol.83, May - June 2020 ISSN: 0193-4120 Page No. 11117 – 11121.

[18] Nallamala, S.H., et al., "A Brief Analysis of Collaborative and Content Based Filtering Algorithms used in Recommender Systems", IOP Conference Series: Materials Science and Engineering, 2020, 981(2), 022008.

[19] Nallamala, S.H., Mishra, P., Koneru, S.V., "Breast cancer detection using machine learning approaches", International Journal of Recent Technology and Engineering, 2019, 7(5), pp. 478–481.