A SURVEY ON PERONALITY PREDICTION USING TWITTER

M.D. Sale[1], Atharva Sahu[2], Rasika V Burde[3], Prachi[4]

[1]Asst. Professor, Dept of Computer Engineering, Sinhgad College of Engineering, Vadgaon, Pune-411041, Maharashtra, India

[2, 3, 4]B.E. Student, Dept of Computer Engineering, Sinhgad College of Engineering, Vadgaon, Pune-411041, Maharashtra, India

Abstract

Now-a-days, Social media has become so popular that you can easily find people of every age group communicating, connecting, and sharing information and ideas with each other. What we are interested in is the personality of users. Personality is a unique trait, which can tell us majorly about a person, for instance, user's pattern of behavior. Thus, predicting personality with help of social media turning out to be a trending research area in computational linguistic. Currently, with the help of social media platforms, more and more people can express their activities, feeling, thoughts, and opinions. Posts, comments and status updates made by users of social media can reveal personal information. Here, Twitter, a huge social media platform, comes into picture. In this paper, we are going to use this bulk of publicly available information to help predict a user's personality. We will describe the type of data collected, our methods of analysis, and the machine learning techniques that allow us to successfully predict personality, and jobs/ professions that are most suitable to a user according to his personality.

Keywords

Social Media, Personality Traits, Machine Learning, LSTM, BERT base, NLP

I. INTRODUCTION

Social networking is booming at a massive rate, with millions of people posting messages or tweets, or photos within a minute. According to a recent survey, Twitter alone has 353 million monthly active users worldwide, with 9,281 tweets sent within a second and 500 million tweets sent per day. Users reveal a lot about themselves while creating their social media profile such as, whom they want to share posts with, what they share. A lot about user's personalities can be exploited from their profile through their posts, status updates, interests.

Personality is the characteristic patterns of thoughts, feelings, and behaviors that make a person unique. Personality affects decisions-making and behavioral activities of person. Traditionally, personality assessment has always been performed by psychology experts, with the help of interviews or self-reports but today technological advancement has provided us a platform with the help of which we can bridge the gap between personality research and social media.

Going through Previous works done in this field, we saw that the information in users' Facebook profiles is not an "idealized" version of themselves, but a reflection of their actual personalities [1]. Making an analogy, we hypothesized other social media platform users to do the same, and Twitter, being one of the most popular and having a head-count of 188 million daily users, became an ideal platform for the study. The process for predicting user's personality goes like this.

1.1 Web scraping User's data and Preprocessing:

Twitter allows developers to scrap publicly available data through a set of private keys i.e., Twitter API keys. These keys allow developer to request data on behalf of App's owner.

Tweepy is a Python library for accessing the Twitter API. The last 50 tweets of user are extracted, and then preprocessed with the help of regex.

Fig 1: Data Extraction

1.2 Natural Language Processing (NLP):

The two primary model used for Natural Language Processing in this system are LSTM and BERT base uncased. The values of the dataset i.e., pre-existing MBTI dataset and the inputs from the CSV file is processed by passing data through these models

Fig 2: Natural language Processing

1.3 Myers- Briggs Type indicator (MBTI):

The MBTI is one of the most popular personality tests in the world. The Myers Briggs Type Indicator (or MBTI for short) is a personality type system that divides everyone into 16 distinct personality types across 4 axes: Introversion (I) – Extroversion (E); Intuition (N) – Sensing (S); Thinking (T) – Feeling (F); Judging (J) – Perceiving (P). The processed data will be analysed through each of personality type indicator and most favourable out of 16 personality type will be allocated to the user.

Fig 3: MBTI Personality Types

II. Literature Survey

Using Neural Network and Natural Language Processing to predict personality and professions suitable according to the proposed personality. The paper proposes a survey that is based on the implementation of neural networks and natural language processing in predicting personality. The proposed system can exploit user's data and reveal user's personality more accurately. people use social media to freely express themselves on issues concerning their lives and family well beings, psychology, their interaction with society and environment, and politic. individual behavior and personality can be characterized using these expressions [2],[4]. Previous personality analysis techniques using RNN and LSTM is compared with pre-trained language model BERT (Bidirectional Encoder Representations from Transformers) [3].

BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. It explains a new language representation model that can be fine-tuned with just one additional output layer to create a state-of-the-art model which can outperform the other models that are used for natural language processing [5].

In Myers-Briggs Personality Classification and Personality-Specific Language Generation Using Pre-Trained Language Models, the use of pre-trained model to predict MBTI personality types based on scrapped texts is examined. Personality classification is complex because of countless factor involver and even a human being might not classify accurately a personality based on a text. However, pre-trained language models might be able to pick up on various subtleties in how different personality types use language [6]. Neural Networks in Predicting Myers Brigg Personality Type from Writing Style, the paper puts forward use of neural network for text base prediction of MBTI personality types. The paper also proves the reliability and advantages neural network by comparing it with other traditional models [7].

Table 1: Comparison Table

Paper Title	Year	Seed Idea
Facebook Profiles Reflect Actual Personality, Not Self-Idealization	2010	-People express their original thoughts on social media instead of an idealized version of themselves
Personality traits recognition on social network—Facebook	2013	-Correlation between user's personality and social media is strong
The development and psychometric properties of LIWC2015	2015	-the work of personality extraction from the text.
Neural Networks in Predicting Myers Brigg Personality Type From Writing Style	2017	-LSTM with other model gives more accuracy. LSTM has potential of More information to be encoded in parameters
BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding	2019	-Natural language processing with BERT obtains state-of-the-art results.
Myers-Briggs Personality Classification and Personality-Specific Language Generation Using Pre-trained Language Models	2019	-Larger and cleaner data sets will increase the accuracy of BERT -BERT model for prediction provide higher accuracy than other models.
Text based personality prediction from multiple social media data sources using pre-trained language model and model averaging	2021	-Deep learning approach with BERT outperforms most personality model -Addition of other NLP statistical feature with BERT can increase model performance.

III. CONCLUSIONS

Using the Bert base uncased and LST for predicting personality is more accurate compared to other models. BERT and LSTM prove to be the most efficient and feasible model in predicting the MBTI personality type. **BERT** is pre-**trained** on a large corpus of unlabeled text including the entire Wikipedia (that's 2,500 million words!) and Book Corpus (800 million words) and natural language processing using BERT already have huge advantage over other models. The proposed system will produce efficient and accurate personality type and professions according to their type.

IV. REFERENCES

[1] M. Back, J. Stopfer, S. Vazire, S. Gaddis, S. Schmukle, B. Egloff, and S. Gosling. Facebook Profiles Reflect Actual Personality, Not Self-Idealization. Psychological Science, 21(3):372, 2010.

[2] J. W. Pennebaker, R. L. Boyd, K. Jordan, and K. Blackburn, "The development and psychometric properties of LIWC2015," Tech. Rep., 2015.

[3] Christian, H., Suhartono, D., Chowanda, A. *et al.* Text based personality prediction from multiple social media data sources using pre-trained language model and model averaging. *J Big Data* **8,** 68 (2021). https://doi.org/10.1186/s40537-021-00459-1

[4]Alam F, Stepanov EA, Riccardi G. Personality traits recognition on social network—Facebook. AAAI Workshop— Technical Report, WS-13-01, 2013. pp 6–9.

[5] BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding Jacob Devlin Ming-Wei Chang Kenton Lee Kristina Toutanova, Google AI Language

[6] Myers-Briggs Personality Classification and Personality-Specific Language Generation Using Pre-trained Language Models Sedrick Scott Keh, The Hong Kong University of Science and Technology

[7] Neural Networks in Predicting Myers Brigg Personality Type From Writing Style Anthony, Stanford University