Myers-Briggs Personality Classification

Using Post Content to Determine Personality Type

Chris O'Brien
Bredesen Center
University of Tennessee
Knoxville, TN USA
cobrien8@vols.utk.edu

Carrie Sanford
Bredesen Center
University of Tennessee
Knoxville, TN USA
csanfor5@vols.utk.edu

Jay Pike
Bredesen Center
University of Tennessee
Knoxville, TN USA
jpike1@vols.utk.edu

INTRODUCTION

The Myers-Briggs Type Indicator (MBTI) is a system that is used to assign a personality type to an individual. Psychoanalyst Carl Jung hypothesized that differences in behavior are not random, but due to the different way people use their perception and judgement, from which the MBTI personality types are derived [1]**.** Typically, these personality types are determined through answering a series of situational or relational questions in which the responses are used for classification. The goal of MBTI is it use an individual's tendencies and preferences to cluster them into predefined categories. There are four initial dimensions: 1) Introversion (I) – Extroversion (E), 2) Intuition (N) – Sensing (S), 3) Thinking (T) – Feeling (F), and 4) Judging (J) – Perceiving (P). The sum of a person's preference for each one of the dimensions represents their individual personality. According to the MBTI system, there are 16 personality types: INFP, INTJ, INFJ, INTP, ENFP, ENTJ, ENTP, ENFJ, ISFJ, ISFP, ISTJ, ISTP, ESFJ, ESFP, ESTJ, and ESTP. For example, the MBTI Foundation describes someone with the ISTP personality as "Tolerant and flexible, quiet observers until a problem appears, then act quickly to find workable solutions". The purpose behind the MBTI as described by the parent company is to "help people understand personality differences in the general population" with a goal of allowing the person to better interact with other people and the world [2]. An example of how the MBTI system could be applied to the real world is through using personality types in the formation of collaborative groups. The idea would be to facilitate a group of compatible and complementary individuals for maximum workplace efficiency. There are four distinct forms of the MBTI in use: Form M containing 93 questions (computer-based or self-scoring), Step II Form Q containing 144 questions, and Step III containing 222 questions. The standard MBTI is reported following Form M evaluation, while Step II and III provide more in-depth analyses, including five facets per each of the four initial dimensions [1]. A successful machine learning model would significantly reduce user time needed to determine their MBTI. This model could then also be expanded to include the five facets within each dimension, should training data be available. Personality type has been predicted using text previously using Reddit posts, comments, and flairs with 82% accuracy with Linguistic Inquiry and Word Count (LIWC) for linguistic feature extraction [4], tweets with 88% accuracy with LIWC, as well as EmoSenticNet and ConceptNet [5], social media posts with 26% accuracy (Naïve Bayes), 33% accuracy (regularized SVM), and 38% accuracy (deep learning) [6], essays and social media posts using Mairesse, SenticNet, NRC Emotion Lexicon, VAD Lexicon and BERT, Albert , and Roberts models for maximum accuracy of 60.6% on essays and 77.1% on social media posts [7]. This work seeks to predict MBTI type with higher accuracy than previous work. The proposed work will be testing the hypothesis that a person's MBTI can be predicted from their post history on a chat forum.

DATA

The dataset was obtained through Kaggle [3] and contains MBTI type for 8675 people. Each MBTI type is matched to the person's last 50 posts on the website PersonalityCafe which is self-described as a "forum community dedicated to all ranges of personality types and people." The users 50 last post are separated by '|||'. The posts range anywhere from strictly text to hyperlinks to other sites to GIFs.

METHODS

The dataset will first be cleaned and analyzed using standard data exploration techniques. Data exploration techniques may include but are not limited to the following: personality type distribution, post length, and separation of initial dimensions ie Introversion/Extroversion, Intuition/Sensing, etc. Raw data will be visualized. Once data has been cleaned and visualized, the text data will undergo ML feature extraction through vectorization and will be split into training (70%), validation (15%), and testing (15%) sets. Supervised machine learning (ML) algorithms will be used for prediction of personality type which provides an n=16 classification problem. Furthermore, data that was separated out by dimension will be used to classify, for example if a person is an introvert or extrovert. The latter problem will be an n=2 classification problem repeated 4 times. The baseline ML models used will be both Logistic Regression and Naïve Bayes. From there the following algorithms, will be implemented: Random Forest, Linear Support Vector Machine, and XGBBoost. Each model will be compared using precision, recall, f-1 score and accuracy. In addition to comparing the results of the different models implemented, we will use the Kaggle submitted code and find a peak accuracy to compare our findings to.

CONCLUSION

The preferred takeaway from this project is a machine learning model able to predict the personality type of an individual based off their forum posts. If a reliable classifier can be constructed, then the benefits described in the conclusion of superior group formation and customized interfaces may be researched further. However, if poor classification results are found, then further analysis will be required to understand the degree of correlation between personality forum data. Poor classification results will nonetheless provide a baseline for future efforts, and they will aid in determining new techniques for future efforts of personality classification.

CITATIONS

1. Anon. MBTI® basics. Retrieved November 5, 2021 from https://www.myersbriggs.org/my-mbti-personality-type/mbti-basics/
2. Anon. MBTI Facts. Retrieved November 5, 2021 from https://www.themyersbriggs.com/en-US/Support/MBTI-Facts
3. Mitchell J. 2017. (MBTI) myers-briggs personality type dataset. (September 2017). Retrieved November 5, 2021 from https://www.kaggle.com/datasnaek/mbti-type
4. Gjurković, M. and Šnajder, J., 2018, June. Reddit: A gold mine for personality prediction. In *Proceedings of the Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media* (pp. 87-97).
5. Bharadwaj, S., Sridhar, S., Choudhary, R. and Srinath, R., 2018, September. persona traits identification based on myers-briggs type indicator (MBTI)-a text classification approach. In *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (pp. 1076-1082). IEEE.
6. Cui, B. and Qi, C., 2017. Survey analysis of machine learning methods for natural language processing for MBTI Personality Type Prediction.
7. Khan, A.S., Ahmad, H., Asghar, M.Z., Saddozai, F.K., Arif, A. and Khalid, H.A., 2020. Personality Classification from Online Text using Machine Learning Approach. *International Journal of Advanced Computer Science and Applications*, *11*(3).