Predicting Myers-Briggs type from conversational writing

Jun Chen[*]
Department of Statistics
University of California, Berkeley
Berkeley, CA 94720
athenachen@berkeley.edu

Abstract

Using data from the Personality Cafe forum, dedicated to discussion of the Myers-Briggs Type Indicator (MBTI), we classify users into one of four groups based on the primary or auxiliary extraverted function of their MBTI type. We extract features from the data using the natural language processing tool UDPipe and build our classifiers using the number of occurrences of the most frequently used words for each type in each user's 50 most recent forum posts as features. Our multinomial logistic regression model yields a prediction accuracy of 0.524, our penalized multinomial logistic regression model an accuracy of 0.522, our random forests model an accuracy of 0.511, and our random forests model sans the least significant features determined by the penalized model an accuracy of 0.512.

1 Introduction

The Myers-Briggs Type Indicator (MBTI) is a personality inventory that has gained a notable reputation across the world. It was developed by Katharine Cooks Briggs and Isabel Briggs Myers during the early to mid 20th century based on the work of psychologist Carl Jung [1].

The assessment assigns takers one of sixteen personality types, each composed of four letters. The first letter describes whether the individual prefers to focus on the external world (Extraverted) or their own internal world (Introverted), the second whether they focus on facts and details (Sensing) or meanings and interpretations of information (iNtuitive), the third whether they make decisions based on logic (Thinking) or the values of the people involved (Feeling), and the fourth whether they prefer to live a structured life (Judging) or a more flexible one (Perceiving) [2].

A more advanced interpretation of each personality type involves assigning "cognitive functions" based on the letters of each type, which describe which processes the individual prefers to use externally and internally. In total, there are eight functions: Extraverted/Introverted Sensing (Se/Si), Extraverted/Introverted Intuition (Ne/Ni), Extraverted/Introverted Thinking (Te/Ti), and Extraverted/Introverted Feeling (Fe/Fi). These were taken directly from Jung's proposed personality types [3].

According to Briggs and Myers, each person operates with a primary and an auxiliary function. The first letter (E/I) determines whether their primary function is Extraverted or Introverted—the auxiliary function is always the opposite of it. The last letter (J/P) determines whether their decision-making (T/F) or perceiving (S/N) function is the Extraverted one. For instance: someone with the INFP type would have Fi as their primary function and Ne as their auxiliary function.

[*]Source code for this project located at https://github.com/athenacy/mbti-prediction/.

Preprint. Under review.

With the rise of the Internet, there has developed something of a subculture around this test. Many people find it fun or intriguing to discuss how their type reflects their behavior and values, or which pop culture characters would have which type. Thus, MBTI enthusiasts have created several online spaces for like-minded people to discuss the test and ideas related to it. These include forum sites such as Personality Cafe, meant for open-ended discussion, and Personality Database, a site specifically for users to assign MBTI types to popular characters.

Even though the test itself has little scientific basis, there is still merit to identifying a correlation between one's self-identified type and how one tends to speak in casual conversation. Perhaps knowing one's type can influence the way one acts; in particular, it may subconsciously or consciously pressure one to fulfill their type's stereotypes (ex. Thinking types like science, Feeling types are overly sensitive). On the other hand, one's existing interests and values can also cause one to gravitate towards identifying with a certain type based on those stereotypes.

Using a dataset scraped from Personality Cafe, we aim to build a classifier that predicts a person's MBTI type based on a corpus of conversational text they have written.

2 Exploratory data analysis and feature extraction

The dataset contains 8675 rows and two columns: one listing the MBTI type of each user, and one containing the last 50 posts each user had on the Personality Cafe forum at the time of collection (September 22, 2017, according to the Kaggle page). Some of the longer posts are truncated after around 200 characters or about 35 words.

In many cases, types with only a one-letter difference are similar enough such that it would be difficult to distinguish between them in casual conversation. Thus, instead of classifying between all 16 types, we divide them into 4 groups based on their main or auxiliary Extraverted function, as it is the function used to interact with the external world, including social situations. We create a new column with factors from 0-3: 0 represents the Ne types (ENTP, ENFP, INTP, INFP), 1 represents the Se types (ISTP, ESTP, ESFP, ISFP), 2 represents the Te types (INTJ, ENTJ, ESTJ, ISTJ), and 3 represents the Fe types (INFJ, ENFJ, ESFJ, ISFJ).

There appear to be vastly more Ne types than other types in the dataset, as well as slightly fewer Se types. This difference in proportions should be kept in mind when performing exploratory data analysis and analyzing the results of the classifier.

Table 1: Number of users per type group

Group	Users
Ne	4496
Se	745
Te	1566
Fe	1868

Following some more cleaning—in this case removing unnecessary quotation marks surrounding some entries in the "posts" column—we split the data into a training and test set, where a randomly selected 80% of the data is assigned to the training set and the remaining 20% is assigned to the test set. We perform EDA and model fitting on the training set, while transformations and feature extraction are applied to the whole dataset.

2.1 Number of links per corpus

The first feature we would like to extract is the number of links in each user's collection of posts. It is reasonable to expect a difference between people of certain types in how often they refer to an external source to express certain ideas, as opposed to expressing those ideas in their own words. Table 2 displays the average number of links used per person for each type group.

Se types tend to use links most often and Te types tend to use links least often. This result is justified since Se types are said to focus on existing facts and details, and thus would be more likely to quote sources directly instead of paraphrase them with their own interpretation. Since there appears to be a

Table 2: Average number of links per user per type group

Group	Average number of links
Ne	3.32
Se	3.89
Te	3.08
Fe	3.27

significant difference between the average links per person for these type groups, we will include this variable in the final model as a new column named "link."

2.2 Adjective-to-word ratio

The next few EDA steps involve the natural language processing toolkit UDPipe. This package contains functions to split a body of text into individual lemmas (base forms of words, i.e. "dog" and "dogs" are considered to have the same lemma "dog"), as well as annotate each lemma with its part of speech. Since links interfere with UDPipe's analysis, and we have already extracted the feature we need involving the links, we remove them from each collection of posts. We use the "English-EWT" model since it is trained on data such as weblogs, newsgroups, emails, reviews, and Yahoo! answers, as opposed to more literary sources.

Using the part-of-speech tagging function, we can extract the ratio of adjectives and adverbs to the total number of words for each user. This feature is of interest to us since a greater frequency of adjectives and adverbs tends to indicate more subjective speech, which can be a good indicator of whether the speaker has a more logical or emotional personality.

Table 3: Adjective-to-word ratio per type group

Group	Adjective-to-word ratio
Ne	0.136
Se	0.131
Te	0.131
Fe	0.134

From Table 3, it is clear that there is no significant difference between the adjective ratio between types, as they all match to 2 significant figures. Thus we will not include it as a feature in the final model.

2.3 Keywords

The last feature we would like to extract is the frequency of certain keywords, i.e. the most frequently used lemmas with a significant difference in frequencies for at least one pair of type groups. To find these keywords and how many times each person used them, we simply use the list of lemmas outputted by UDPipe's annotation function for each corpus of posts in the training data.

For each type group, we calculate how many times each lemma was used by all users of that group, then divide these counts by the total number of words across every post made by members of that group. We then filter this list of lemmas, keeping those where the largest relative difference between their largest and smallest per-group frequency is greater than 0.5. This threshold was chosen carefully, as setting this number too high would cause too many common words to be filtered out, leaving behind too many technical words (such as MBTI types, the abbreviations for cognitive functions, and related personality indicators such as Enneagram types). A preliminary list of keywords is obtained by taking the 200 most frequently used lemmas in this filtered list.

Most notably, emoticons and swear words, type and function abbreviations, and words stereotypically associated with certain types such as "feel" and "science" appear in this list. Unsurprisingly, "feel," "love," and ":)" were mentioned most by Fe types and least by Te types, and "science," "smart," and "logic" were mentioned most by Te types and least by Fe types. Profane words are consistently used most by Se types and least by Fe types.

To avoid overfitting the list of keywords to this dataset, we remove all the MBTI-related terms, including type and function abbreviations, as well as "Mbto" (a misspelling of MBTI). We also remove forum-specific terms; namely "PerC" (abbreviation for Personality Cafe), "Tapatalk" (the platform the forum is hosted on), and OP (forum-speak for "original poster"). The final list is 155 words long and can be found in the linked GitHub repository.

Table 4: Frequencies per type group of the ten most common keywords

	Frequency			
Lemma	Ne	Se	Te	Fe
feel	2.56e-03	2.16e-03	1.80e-03	2.97e-03
love	1.93e-03	1.65e-03	1.28e-03	1.99e-03
:)	8.66e-04	7.07e-04	5.30e-04	1.05e-03
yeah	6.82e-04	8.38e-04	5.28e-04	5.75e-04
thank	5.31e-04	4.00e-04	4.61e-04	6.93e-04
stuff	4.12e-04	5.55e-04	3.28e-04	3.50e-04
such	4.32e-04	3.09e-04	4.31e-04	4.65e-04
fun	3.99e-04	5.17e-04	3.26e-04	3.21e-04
music	4.19e-04	4.52e-04	2.83e-04	3.39e-04
however	3.61e-04	2.97e-04	4.60e-04	3.26e-04

Afterwards, we add one column to the full dataset for each of these keywords. These columns contain the number of times the corresponding word is used in each person's posts. Table 5 shows an example of these features.

Table 5: Number of times the first six users mentioned the five most common keywords

Group	feel	love	:)	yeah	thank
Fe	0	0	0	0	0
Ne	0	1	5	0	0
Ne	2	1	7	0	2
Te	1	0	0	1	0
Te	0	0	0	0	0
Te	1	1	0	0	0

3 Model fitting and prediction results

3.1 Multinomial logistic model

We first fit a multinomial logistic regression model to the dataset containing all our desired features, predicting each user's type group from all other variables. We use this model since it produces consistent results, due to the algorithm being purely mathematical, and can serve as a good estimate of the prediction accuracy using other, more random methods. A prediction accuracy of 0.524 is obtained when this model is run on the test set.

3.2 Penalized multinomial logistic model

The next method we use is a penalized multinomial logistic regression model. Similar to LASSO for linear regression, this method uses a grouped L1 penalty to shrink the coefficients of the least useful features to zero, making it useful for feature selection. Before fitting the model, we use cross-validation to tune the magnitude of the penalty term.

We obtain a prediction accuracy of 0.522. The model sets the coefficients of four keywords to zero: "!!!," "correct," "statement," and "dude," as well as the coefficient for the number of links per corpus.

3.3 Random forests model

Finally, we fit a random forests model to both the original dataset and the dataset without the features whose coefficients were set to zero in the penalized multinomial logistic regression step. We obtain a prediction accuracy of 0.511 for the original dataset and 0.512 for the reduced dataset. This suggests that removing the coefficients offered a small improvement in prediction accuracy, though due to the random nature of the fitting method and the small magnitude of the difference between the two accuracies, we cannot yet say whether the improvement is significant.

4 Limitations

The dataset we used is from a forum dedicated to the MBTI test. This almost certainly biases the topics that are discussed on the site, and thus the collections of text in the data; for instance, MBTI types are rarely brought up outside interest groups dedicated to the subject. General psychology topics are also more likely to be found on the forum compared to most real-world conversations.

As mentioned previously, posts that are too long are truncated, which affects the extraction of features such as word count, as well as frequency of certain words or parts of speech. We were unable to use the total word count of each post or each sentence as a feature due to this. Also, adjectives and adverbs tend to appear earlier in sentences than nouns and verbs, so cutting sentences off in the middle may cause adjective-to-word ratios for a particular person to be calculated as larger than they actually are.

Finally, according to a comment on the dataset's Kaggle page, the type frequencies of the forum data may not match that of the world population. Certain types being over- or underrepresented in the forum data compared to the population may affect classification accuracy.

Table 6 contains the frequency of each type in the dataset, compared to the frequency of each type in the population, based on data compiled from MBTI tests administered between 1972 and 2002 [5]. Since the population data is somewhat outdated, it may not be accurate to the time the forum data was collected, but it does serve as a ballpark estimate of population type frequencies.

Table 6: Percentage frequency of each type in the dataset and in the population

	Percentage	
Type	Data	Population
ENFJ	2.2	2.5
ENFP	7.8	8.1
ENTJ	2.7	1.8
ENTP	7.9	3.2
ESFJ	0.5	12.3
ESFP	0.6	8.5
ESTJ	0.4	8.7
ESTP	1.0	4.3
INFJ	16.9	1.5
INFP	21.1	4.4
INTJ	12.6	2.1
INTP	15.0	3.3
ISFJ	1.9	13.8
ISFP	3.1	8.8
ISTJ	2.4	11.2
ISTP	3.9	5.4

5 Conclusion

Our methods for predicting the main extraverted function of an individual's MBTI type based on keywords found in conversational text written by them are somewhat effective, having marginally better accuracy than random guessing (prediction accuracy of 0.506 if one guesses group 0 every time on the test data). These results are notably influenced by the specific split between train and test data, so it is hard to say whether they are definitive. We have yet to see how well these classifiers perform on data gathered outside the Personality Cafe forum.

Broader Impact

The MBTI assessment is largely regarded as pseudoscience. Despite this, many psychologists and career counselors do utilize their clients' MBTI types in their work, and the work outlined in this paper may benefit them. However, it may also be used to further stereotypes relating to each MBTI type, potentially hurting said clients by pigeonholing them into certain treatments or opportunities that may not be suitable for them. This work may also benefit those who analyze MBTI typings for fun, as it may help them gain insight to how their manners of speech are reflected by their own type or help them assign a type to fictional characters.

Acknowledgments and Disclosure of Funding

We are grateful to Nikita Zhivotovskiy and Saptarshi Chakraborty for their advice and guidance in completing this project.

References

[1] The Myers-Briggs Company. (2019). *The history of the MBTI® assessment.* The Myers-Briggs Company. https://eu.themyersbriggs.com/en/tools/mbti/myers-briggs-history

[2] The Myers & Briggs Foundation. (2003). *MBTI® basics.* The Myers & Briggs Foundation. https://www.myersbriggs.org/my-mbti-personality-type/mbti-basics/

[3] Ross, F. (2015, August 12). *Carl Jung personality types.* The SAP; The Society of Analytical Psychology. https://www.thesap.org.uk/articles-on-jungian-psychology-2/about-analysis-and-therapy/typology/

[4] Jolly, M. (2017, September 22). *(MBTI) Myers-Briggs personality type dataset* [Data set]. https://www.kaggle.com/datasets/datasnaek/mbti-type

[5] The Myers & Briggs Foundation. (2003). *How frequent is my type.* The Myers & Briggs Foundation. https://www.myersbriggs.org/my-mbti-personality-type/my-mbti-results/how-frequent-is-my-type.htm