Evaluating the MBTI Personality Construct Using Text Data

Prepared by Benjamin Olsen

***Abstract* - This project is an effort to test the hypothesis that the Myers Briggs Type Indicator can appropriately be used to classify text. Features are extracted from the labeled text by preparing a bag-of-words vector space model. The extracted features are then used to train 4 separate support vector machines for binary classification in each personality axis of the MBTI (introversion vs. extroversion, intuition vs. sensing, thinking vs. feeling, and judging vs. perceiving) . This is done to mirror the original scoring method by which the MBTI test proposes to describe personality. An average precision of 0.841 is achieved across each axis, and the distribution of the classification probabilities suggests that each axis is describing two distinct populations within that axis. When a bias in the text data is removed however, an average precision of 0.622 is achieved and the distribution of the classification probabilities suggest that each axis contains only one population. While this indicates that the MBTI is not appropriate for classifying text data in unbiased situations, it also implies a relationship between the context of the text data and the characteristics of the data.**

I. INTRODUCTION

This project is an effort test the hypothesis that the Myers Briggs Type Indicator can appropriately be used to classify text. The Myers Briggs Type Indicator (or MBTI) is an "introspective self-report questionnaire with the purpose of indicating differing psychological preferences in how people perceive the world around them and make decisions" according to Wikipedia. While the professional field of psychology has discarded this test due to its poor retestibility results, the test's usefulness has yet to be evaluated in the context of online behavior. This project explores whether the MBTI can be used to classify text data, and if MBTI is "appropriate" for this use in that the MBTI personality construct clearly manifests itself in text features generated by individuals online.

II. BACKGROUND DETAILS

The Myers-Briggs Type Indicator test is a series of questions designed to score individuals along four separate axes. According to their answers, test-takers receive a score for each axis, which represents certain personality aspects. The axis are:

- Introversion (I) – Extroversion (E)
- Intuition (N) – Sensing (S)
- Thinking (T) – Feeling (F)
- Judging (J) – Perceiving (P)

Each axis end is assumedly opposite in characteristic to its other end of the axis, and each axis is split down the center. If a person scores to one side of the center, they are assigned the attribute for the side they scored closest to. For instance, on the I (Introvert) to E (Extrovert) scale, someone can score closer to I than to E, and they are assigned the letter 'I' for that scale, indicating that they are more introverted than extroverted. Once the scores from all the axis are completed, the test-taker receives a string of letters that represents their personality traits. For example, an INFP is considered to be introverted, intuitive, feeling, and perceiving.

Figure 1. The 4 MBTI axes.

The Myers-Briggs Type Indicator (MBTI) test faces scrutiny due to poor retest. What is often found is that, upon retaking the test, test-takers often receive slightly varying scores. With enough scores collected, it can be observed that they tend to be normally distributed around the center of the axis, suggesting the scores can not reasonably be divided into two populations (much less polar opposites). This highlights the dynamic nature of personality and calls into question the validity of the MBTI as a personality construct. However, the MBTI personality construct is quite popular online, meaning that while other more clinical methods may be much more reliable, the MBTI has much more readily available data.

III. MATERIALS

The data set in question for this analysis contains over 8600 rows, in each of which is a person's "Type" (This person's 4 letter MBTI code/type), and a section of each of the last 50 things they have posted (Each entry separated by "|||" pipe characters) (*Figure 2)*.

This data is collected through the PersonalityCafe forum, and contains a large selection of people and their MBTI personality type, as well as what they have written. Further information on this data set can be found at:
<https://www.kaggle.com/datasnaek/mbti-type>

	type	posts
0	INFJ	'http://www.youtube.com/watch?v=qsXHcwe3krw|||...
1	ENTP	'I'm finding the lack of me in these posts ver...
2	INTP	'Good one _____ https://www.youtube.com/wat...
3	INTJ	'Dear INTP, I enjoyed our conversation the o...
4	ENTJ	'You're fired.|||That's another silly misconce...

Figure 2. Sample of the raw labeled data.

IV. EXPERIMENTAL DESIGN

1. Data Wrangling

For this particular data set, each document is a collection of 50 comments collected from one user, with each comment separated by three pipe characters (|||), so to maintain the overall size of the dataset the pipe characters were removed but the comments were kept together in one long document, effectively maintaining the original size of the data set with full features.

The next steps include removing escaping HTML characters, removing hyperlinks, expanding contractions, removing digits, removing all punctuation, and removing stop-words (frequently occuring words with no intrinsic value, such as 'the', 'and', 'a').

Another set of the text data was created in which all mentions of any of the 16 classes by name removed. My reasoning for this step is to simulate text data collected from a different source. Since the text was collected from a forum dedicated to personality types, the data contains an inherent contextual bias, as the people generating the data might not exhibit the same behavior in a space that is not explicitly denoted for discussion of personality types.

For simplicity's sake, I will henceforth refer to the cleaned documents as Sample 1, and the cleaned documents with class names removed as Sample 2.

2. Feature Extraction

This analysis was conducted using the Bag-of-Words vector-space model, which turns each word appearing in the corpus into a feature by which a machine learning model can be trained. The result is a very sparse feature matrix with *k* features and *n* samples. Each feature is weighted using term frequency–inverse document frequency which normalizes each feature based on a weighting assigned by frequency of appearance. Each document is then converted into a vector, which is essentially a closest fitting line to each feature within a document so that vectors can be compared between classes utilizing cosine similarity (*Figure 3*).

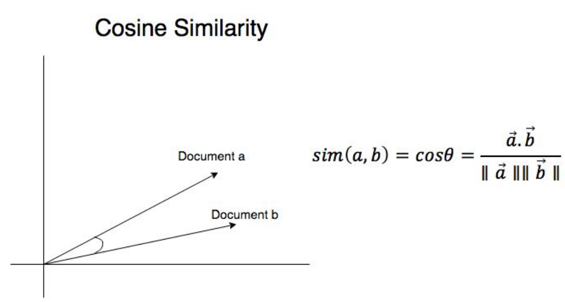

Figure 3. Cosine similarity.

3. Machine Learning

Considering the theoretical MBTI construct (4 separate axes with two categories each), the inherent nature of the classes meant that any given document would share some of its features with at least half of the other classes. For example, all classes starting with ‘E’ would share overlapping features with one another, adding noise a model and confusing each class.

This was the reasoning behind the choice to train four separate binary classifiers; one for each axis. This is effective for reducing the noise in the data by allowing the classifiers to focus on only two hypothetically polar aspects of the data at a time, while also increasing the support size for each class.

There existed a skewed representation of classes in the data (*Figure 4*), which was mitigated by balancing the weighting of the classes while fitting the classifier to the training data.

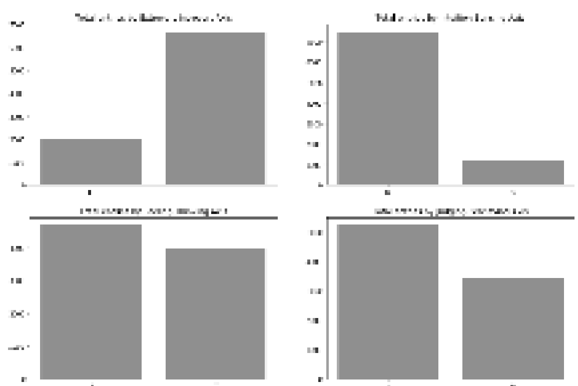

Figure 4. Unequal support for each class within each axis.

Utilizing a Stochastic Gradient Descent Classifier with a logarithmic loss function enables the classifier to predict the probability of a classification given its vectorized text features. Since the probabilities of binary classification are proportional (75% probability of class A infers 25% probability for class B), this approach is effective in simulating the original scoring method in which test scores fall along each axis, but this time the scores represent probabilities of each classification based on vectorized text features classified by a support vector machine.

This approach is necessary to test our hypothesis (the MBTI personality construct can be reliably used to classify the text that a person posts online), the reason being that if they MBTI theoretical structure of personality is to be considered reliable, there must exist a minimum of two populations within each axis. This would offer strong support to the idea that personality can be broken down in such a fashion, and that that structure is visible within text features. A strong bimodal distribution of the classification probabilities between the classes would indicate that the classifier was confident in its classifications, thus supporting the MBTI presumption that people tend to fall towards one end or the other of each axis. Conversely, a normal distribution of scores would indicate that the MBTI is incorrectly attempting to separate one population into two under false assumptions, and thus causing us to accept the null hypothesis.

V. RESULTS

Classification fared well, but there there is a clear difference between Sample 1 and Sample 2 in precision (*Figure 5*). To explore this further, the probability distribution must be investigated.

Precision of SGDClassifier per Sample

Axis	Sample 1	Sample 2
EI	0.842	0.577
NS	0.875	0.579
FT	0.847	0.763
JP	0.799	0.568

Figure 5. Performance of SGDClassifier on each sample by axis.

The initial results on Sample 1 test data, while not perfect, are promising. The lack of support in certain classes is affecting its ability to predict them, but with the balanced weighting there is a reasonable bimodal distribution along each axis, supporting the idea that the axis dichotomy is appropriate for dividing any population of text into two distinct populations due to distinctive text features (*Figure 6*). Pertaining to Sample 1, the results seem to suggest support for using the MBTI as a method of classifying text.

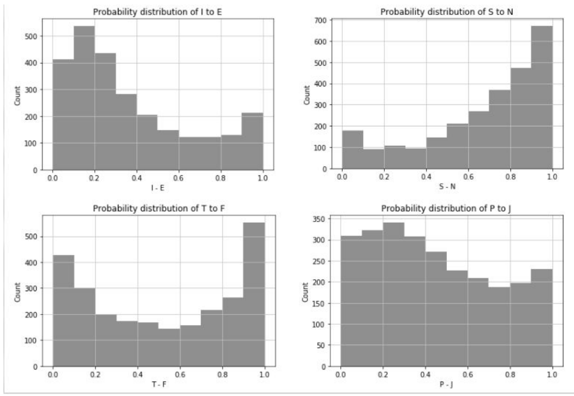

Figure 6. Probability distribution of Sample 1 test data.

However, this success needs to be viewed in context to Sample 2 (recall that Sample 2 had all mentions of any of the 16 classes removed). The test data results after training the classifier with Sample 2 returned are nearly perfect normal distributions for the classification probabilities (*Figure 7*).

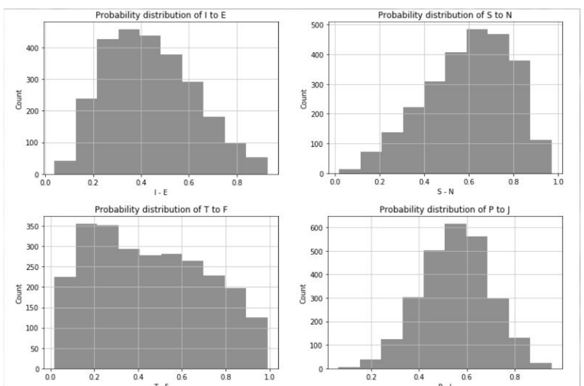

Figure 7. Probability distribution of Sample 2 test data.

VI. DISCUSSION

It would seem that the filtering out of all 16 class references removed one of the best distinctive features across all 8 classes in each of the 4 axis. From this it is apparent that the classifier trained on Sample 1 did not generalize over in terms of predicting power on Sample 2, suggesting that each of the 16 class references when present in the data are the most discriminative features of the personality classes.

From the perspective of this data, it is clear that the context of where the data is being gathered is quite important. This data was gathered from a forum where people discuss their personality types with one another openly. More importantly, they are aware of their personality type and are willing to share about it. Naturally this makes the data contextually reliant: in a place where people are aware of and constantly sharing about their personality types, they are also likely to adhere to certain behavioral patterns. In simpler words, whether or not someone is aware of their "personality type" and how it might influence their behavior seems to affect how well the text they output can be classified utilizing the MBTI.

In terms of machine learning, what these results might imply is that this model wouldn't generalize well to data collected from other areas (where the context is different) in any practical sense.

While the MBTI did not prove to be a practical way to classify text data, it certainly holds lvalue in describing the behavior of a subset of people, as it clearly holds value to a large number of people to the point where knowledge of it might even dictate their behavior.

An obvious application for this model lies in the realm of quantifying impact of a piece of literature on a target audience, which in this case would be people who identify strongly with the MBTI. The only current metric to measure impact is views, shares, and comments. These are strong, 'proof-in-the-pudding' kinds of metrics, but are not available before the content is actually delivered. This makes content delivery a kind of trial and error process to find the right content to deliver to the target audience to get the best response. With this model in particular, a document could actually be scored on the probability that it represents particular personality class elements before even being presented. This model of course only would apply to text from the MBTI perspective.

VII. CONCLUSION

Hopefully then this analysis can serve as a strong reminder of the power of context. Text is something that is readily available in multitudes of observations, and is so often discarded as being too complex to be useful. From what we can see in this applied analysis, context can add a powerful element to what might otherwise be a useless bag of words, and when considered carefully enough can actually add significant value to what otherwise might be considered the noisiest kind of data around.