Analysis of Linguistic Characteristics of Persuasive Speech and Its Predictive Ability of Personality Traits and Social Value Orientation of Speakers

Simon Risman **Harrison Finkelstein-Hynes** **Ada Tur**

Abstract

In this work, we analyze a series of conversations from the CaSiNo Corpus [1] and Persuasion For Good Corpus [2], which involve interactions between two speakers, who must work to reach a set goal using persuasion and problem-solving skills. For each speaker, we train a classification model to predict their personality traits and social value orientation, and we also compare these results with human evaluations and predictions from ChatGPT.

Introduction

A famous quote from Aristotle loosely reads "Character may almost be called the most effective means of persuasion", to mean that one's character and impression they give may be the most important step of their process of persuasion and convincing [3]. However, a more interesting question is can you predict one's character based on their persuasion instead? With the adoption of language models as a means for analyzing human speech, it is easier than ever to make predictions off of such data. But can we target the core values and traits of a human being based solely on a text conversation? Our initial hypothesis is yes - on the basis of text and transcribed oral exchanges, it is entirely possible to estimate one's traits based on how they interact with others, particularly when they have an end goal in their personal interest. To go about testing this hypothesis, we utilize two corpora: the CaSiNo (CampSite Negotiations) Corpus and the Persuasion For Good Corpus, both from ConvoKit [4]. The CaSiNo Corpus is a collection of over 1000 written conversations between two speakers where the goal is to divide some natural resources (3 wood, 3 water, 3 food) for a hypothetical camping trip. The speakers must present what they would like to get out of a deal, and negotiate for the supplies until a final deal is reached and agreed upon by both participants. The Persuasion For Good Corpus (PFG) similarly has around 1000 text conversations between two speakers. However, in this corpus, one participant is assigned the Persuader, and the other the Persuadee. The goal of the Persuader is to convince the Persuadee to donate to some charity. For both corpora, certain personality traits for each speaker are given. The ones we are interested in are their big-five and their social value orientation. Big-five (B5) is a set of personality traits: conscientiousness, extraversion, agreeableness, openness-to-experiences, and emotional stability. Each trait is given a value between 1 and 7, 7 being most conscientious, most agreeable, etc. For simplicity, we classified each score as above or below the mean value for each B5 trait. Social value orientation (SVO) is a binary category; a speaker's intentions are either "pro-social" or "pro-self". With this data, we explore our initial hypothesis.

Related Work

Before testing our prediction, it is first important to acknowledge the efforts made alongside and beyond our work.

A work from Kosan et al. details an effort at utilizing LSTMs to predict personality traits [5]. The paper first details existing studies over a series of social media sites, a wide variety of preprocessing methods are used with a collection of models to predict different categorizations of personality types. The authors then create a collection of 13 words that posses both positive and negative sentiment, and a dataset of over 11 million tweets was collected accordingly, with user data for around 5000 users. Essentially, if a user had tweeted something containing one of the 13 words in a certain time frame, and they met a set of conditions set by the authors, then the remainder of their tweets were also collected. Each user's personality was categorized with the B5 format, though the paper calls it OCEAN

(Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism), though it is easy to see that these formats are analogous. An LSTM and Bi-LSTM were then used to predict values for these categories, and results were listed as Root Mean Square Error on the basis of the preprocessing method used.

A similar work from Metha et al. makes similar attempts, though with different data, preprocessing methods, and models [6]. The work uses psycholinguistics features, in addition to the language model, to predict certain personality features. Using LIWC, SenticNet, NRC Emotion Lexicon, and others, certain words with specific annotations and sentiments were extracted. Readability was also considered for this evaluation. Then the BERT model (and some of it's counterparts, like Albert and Roberta) is used, though only the BERT model results are presented. Compared with more traditional vocabulary-based methods, larger language models were able to attain much higher results, the state of the art for the task, in fact. The paper displays an important example of the importance of psycholinguistics features, as well as the role of large language models in a more contemporary experimental space.

Further, some researchers have found it interesting to learn about how an interviewee's answers to open-ended questions in an interview may be indicative of their personality traits. Jayaratne et al. explore how the Random Forest algorithm used to train a regression model could be utilized for predicting a participants HEXACO traits, HEXACO being another categorization very similar to B5, standing for Honesty-humility, Emotionality, eXtraversion, Agreeableness, Conscientiousness, and Openness. Results using LIWC, Word2Vec, Doc2Vec, TD-IDF, and some other variations, were compared for results with the regression model, with different approaches having better performance for different traits [7]. For instance, TD-IDF was the most efficient in predicting Honesty-humility. In a future study mentioned by the authors, the participants in the interviews themselves had agreed with the predicted personality trait values, with an accuracy of nearly 88%.

Lastly, we found the 2016 work by Farnadi et al. on personality recognition in social media to be informative for our task as well [8]. Datasets of user information and output from popular social media sites like Facebook, Twitter, and YouTube were collected. An interesting strategy used by this work is the utilization of transcripts of YouTube videos, where users can be seen discussing various topics with a level of transparency and openness that provides very interesting data for the analysis. Similar to some of the other discussed related work, this paper uses LIWC, NRC, MRC, and other psycholinguistics databases in order to extract more information about specific words and terms in the text. Then, six regression models were analyzed on this data: single-target, multi-target stacking, multi-target stacking corrected, ensemble of regressor chains, ensemble of regressor chains corrected, and multi-objective random forest. These six models were the state of the art for this task at the time of the publication of the work. These models were used to answer three questions: "Should personality prediction be treated as a multi-label prediction task, or should each trait be identified separately?", "Which predictive features work well across different on-line environment?", "What is the decay in accuracy when porting models trained in one social media environment to another?" The paper answers the questions: Firstly, the difference between univariate and multivariate models was slim, though the multi-target stacking correct and ensemble of regressor chains corrected both performed the best. So, it seems personality prediction as a multi-label prediction task performed the best. Second, though reaching more of a vague conclusion for this question, the authors found that correlations between linguistic and contextual features and personality traits was tough to generalize, and more important information was extracted from experimentation with the size of the feature space, and its effect on model efficiency, over performance. Third, after conducting six cross-media learning experiments, the authors found that adding more training examples to a model from other social media sites didn't have a significant impact on its performance; most of the results across social media sites remain the same, with minimal fluctuation. Though the contextual features and information from different sites may differ based on the traditional forms of information present on each site, there was overall not a large variation in the models' overall performance. As a result, the work displays important advancements in analyzing online information from what users unprompted choose to distribute on the internet, and how it can be used for personality trait recognition.

Method

To investigate the classification of B5 traits we combined the PFG and the CaSiNo corpora, filtering out any speakers without the appropriate associated metadata. As the B5 traits were on different scales for CaSiNo and PFG, we scaled the CaSiNo 1-7 to the PFG 1-5 ranking.

In initial research we found that the SVC classifier was performing the highest, so we utilized it to perform classification for each B5 trait. Additionally our initial stages of experimentation found that normal preprocessing techniques - stop-word removal, lowercase casting, lemmatization, etc. - made the classifiers perform significantly worse. This is likely because the text itself was overall not very rich. The exchanges were rather generic, especially in the CaSiNo corpus. We therefore performed the classification on unprocessed texts to utilize the maximum number of features possible, even including the emojis that appeared frequently in the CaSiNo corpus. We created 5 classifiers, one for each B5 trait.

The SVO for speakers was only annotated in the CaSiNo corpus, so we could only use one of the corpora for this section. Every classifier we attempted yielded over-prediction of the prosocial category, but the Logistic Regression method was the most balanced.

Once these classifiers had been trained we created a survey with 20 randomly chosen speakers' conversations, given as they were given to the classifier in order to establish a human benchmark for this task. That is, without the corresponding responses and without pre-processing. We considered providing the human and LLM evaluators with the corresponding responses to provide more context, but in the interests of a fair experiment we decided to only provide them with one side. The addition of the corresponding discourse would have introduced too many confounding variables that the classifiers did not have access to - they would no longer just be looking at lexical choices, but at strategies, higher level goals, interpersonal dynamics, and more that we could not account for.

In the interest of exploring the capabilities of LLMs in lexical classification of personality traits, we also used the survey questions given to the human evaluators as prompts to a GPT-3.5 model.

The collected results were analyzed for accuracy to create reference points on performance in this task and better help us evaluate out classifiers.

Results

The results of the classifier are displayed in Tables 1-6 for each of the five categories in B5 and for SVO. Table 1 details the results for SVO, where the classifier was efficient in classifying pro-self correctly, but not pro-social. Tables 2-6 shows the results for extraversion, agreeableness, conscientiousness, neuroticism, and openness, respectively, where the classifier had similar results for above and below mean, and between each category. The scores for each category were consistently between 0.5 and 0.8, showing that the classifier performed better than a randomized coin toss, but not as well as we had hoped for.

	Precision	**Recall**	**F-1 Score**
Pro-Self	0.01	0.50	0.03
Pro-Social	0.99	0.60	0.74

Table 1: Social Value Orientation

	Precision	**Recall**	**F-1 Score**
Above Mean	0.62	0.75	0.68
Below Mean	0.64	0.50	0.56

Table 2: Extraversion

	Precision	**Recall**	**F-1 Score**
Above Mean	0.61	0.77	0.68
Below Mean	0.61	0.42	0.50

Table 3: Agreeableness

	Precision	Recall	F-1 Score
Above Mean	0.57	0.98	0.72
Below Mean	0.62	0.04	0.08

Table 4: Conscientiousness

	Precision	Recall	F-1 Score
Above Mean	0.69	0.53	0.60
Below Mean	0.60	0.75	0.67

Table 5: Neuroticism

	Precision	Recall	F-1 Score
Above Mean	0.50	0.71	0.59
Below Mean	0.59	0.37	0.45

Table 6: Openness

We also compared our results with human evaluation and ChatGPT predictions. For the human evaluation portion, we compiled the first 20 test examples into a Google Form, asking participants to read the speaker's written messages and rank their B5 and SVO as "High"/"Low" and "Pro-Social"/"Pro-Self", respectively. We had a total of 4 participants filling out the form, and Table 7 details the average accuracies for the human evaluation task, where each accuracy is the overall average for the accuracies of each of the 20 conversations' speakers' traits. The human evaluation results show results better than a random coin toss, though not better than our classifier.

Cons	Extr	Agre	Neur	Open	SVO
0.51	0.49	0.58	0.56	0.56	0.46

Table 7: Average Accuracies for Human Evaluation

We also experimented with the performance of large language models on this task, using ChatGPT to compare our results with a stronger model that wasn't trained specifically for this task. Using the prompt:

I am going to give you a series of text that is a collection of messages from one speaker in a conversation from the CaSiNo corpus. Given this set of messages, you will give me a value for five personality traits: conscientiousness, agreeableness, extraversion, emotional stability, and openness to experiences. For each of these five categories, you will return either "High" or "Low". Additionally, you will give me their social value orientation, i.e. either "Pro-Self" or "Pro-Social"

The results for this experimentation are shown in Table 8, where one can see that the performance of ChatGPT was far below the performance of human evaluation, our classifier, and a randomized coin toss. These results give insight into the difficulties a language model can have with predicting personality given textual information; clearly, it isn't as simple of a process as one would expect.

Cons	Extr	Agre	Neur	Open	SVO
0.23	0.18	0.3	0.33	0.33	0.25

Table 8: ChatGPT Accuracies

Discussion and Conclusion

Overall most of our classifiers, with the exception of conscientiousness and social value orientation had f-1 scores of between 55 and 70, which, given the only-text nature of the input data is an impressive result. At its root this task is in itself incredibly difficult. Other classification tasks, like spam filtering or topic recognition require far less data to make distinctions, but when attempting to categorize human personality we recognize that a vast array of data could be and is necessary. In [8], which also tackles a computational approach to personality recognition, the data sets gathered have a much richer set of metadata, gathered from Facebook, YouTube, and Twitter. This contained everything from Facebook's interests and activities section, tags in other people's posts, emotions expressed on posts, and a host of other very pertinent user preferences. While text and speech can tell us a lot about an individual, their B5 personality trait scores are much more difficult to predict. Additionally, besides linear classifiers the some of the other papers used more complex models, such as random forest classifiers and LSTM based prediction models. These models have the potential to identify more latent features that could aid in the recognition of personality traits and social orientation.

Our hypothesis was proven incorrect. Text by itself, even when restricted to the domain of negotiation and persuasion does not contain enough information to classify personality traits or social orientation. While additional labels could enhance this classification capability, text and transcribed speech by itself is insufficient.

Statement of Contributions

The contributions of each project member are detailed below:

Simon Risman: Wrote up Methodology and Discussion sections of the paper, aided in design and construction of classifiers, conducted analysis of social value orientation.

Harrison Finkelstein Hynes: Conducted majority of classification and coding tasks, delved deeply into ConvoKit package to extract maximal information, experimented with variety of models and techniques. Explored multuiple potential avenues for expanding the experiment.

Ada Tur: Wrote up majority of final paper (all portions not mentioned in Simon's description), conducted human evaluation preparation, data collection, and analysis, conducted large language model preparation, data collection, and analysis.

Collaborative: Initial project ideas, proposals, and initial stages of data experimentation were done collaboratively.

References

[1] Chawla, Kushal, et al. "CaSiNo: A corpus of campsite negotiation dialogues for automatic negotiation systems." arXiv preprint arXiv:2103.15721 (2021).

[2] Wang, Xuewei, et al. "Persuasion for good: Towards a personalized persuasive dialogue system for social good." arXiv preprint arXiv:1906.06725 (2019).

[3] Aristotle. The Rhetoric of Aristotle : a Translation. Cambridge [Cambridgeshire] :The University Press, 1909.

[4] Chang, Jonathan P., et al. "Convokit: A toolkit for the analysis of conversations." arXiv preprint arXiv:2005.04246 (2020).

[5] Kosan, Muhammed Ali, Hacer Karacan, and Burcu A. Urgen. "Predicting personality traits with semantic structures and LSTM-based neural networks." Alexandria Engineering Journal 61.10 (2022): 8007-8025.

[6] Mehta, Yash, et al. "Bottom-up and top-down: Predicting personality with psycholinguistic and language model features." 2020 IEEE International Conference on Data Mining (ICDM). IEEE, 2020.

[7] Jayaratne, Madhura, and Buddhi Jayatilleke. "Predicting personality using answers to open-ended interview questions." IEEE Access 8 (2020): 115345-115355.

[8] Farnadi, Golnoosh, et al. "Computational personality recognition in social media." User modeling and user-adapted interaction 26 (2016): 109-142.