

A Multivariate Regression Approach to Personality Impression Recognition of Vloggers

Golnoosh Farnadi*
Dept. of Applied Mathematics
& Computer Science
Ghent University, Belgium
Golnoosh.Farnadi@ugent.be

Shanu Sushmita
Center for Data Science,
Institute of Technology
University of Washington,
Tacoma
shanusushmita@gmail.com

Geetha Sitaraman
Center for Data Science,
Institute of Technology
University of Washington,
Tacoma
sgeetha@uw.edu

Nhat Ton
Center for Data Science,
Institute of Technology
University of Washington,
Tacoma
nhatton7@uw.edu

Martine De Cock
Center for Data Science,
Institute of Technology
University of Washington,
Tacoma
mdecock@uw.edu

Sergio Davalos
Milgard School of Business
University of Washington,
Tacoma
sergiod@uw.edu

ABSTRACT

Research in psychology has suggested that behavior of individuals can be explained to a great extent by their underlying personality traits. In this paper, we focus on predicting how the personality of YouTube video bloggers is perceived by their viewers. Our approach to personality recognition is multimodal in the sense that we use audio-video features, as well as textual (emotional and linguistic) features extracted from the transcripts of vlogs. Based on these features, we predict the extent to which the video blogger is perceived to exhibit each of the traits of the Big Five personality model. In addition, we explore 5 multivariate regression techniques and contrast them with a single target approach for predicting personality impression scores. All 6 algorithms are able to outperform the average baseline model for all 5 personality traits on a dataset of 404 YouTube videos. This is interesting because previously published methods for the same dataset show an improvement over the baseline for the majority of personality traits, but not for all simultaneously.

Categories and Subject Descriptors

I.2.6 [Computing Methodologies]: ARTIFICIAL INTELLIGENCE—*Learning*

Keywords

Multimodal personality recognition; Big Five personality model; personality impression prediction; multivariate regression; YouTube vlog data.

*Also at Katholieke Universiteit Leuven, Belgium

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

WCPR 2014 Orlando, Florida USA

Copyright 2014 ACM 978-1-4503-3129-6/14/11 ...\$15.00.

<http://dx.doi.org/10.1145/2659522.2659526>.

1. INTRODUCTION AND BACKGROUND

Research in psychology has suggested that behavior and preferences of individuals can be explained to a great extent by underlying psychological constructs (or so called personality traits). Knowledge about an individual's personality can allow us to make predictions about preferences across contexts and environments, and enhance recommendation systems [9, 16]. A traditional approach to identify an individual's personality is through a survey. A well known test is the Big Five inventory questionnaire [11] which asks participants to evaluate on a 5 point scale how well their personality matches a series of descriptions for each of 5 personality traits (Extraversion, Agreeableness, Emotional Stability, Openness to experience and Conscientiousness).

More recently, studies have been published about automatically recognizing the personality traits of users of social media platforms based on their user generated content. For example, Oberlander et al. have analyzed blog posts by extracting n-gram features to automatically classify the personalities of 70 bloggers [15], while Gill et al. have analyzed the word usage of 2,400 bloggers to find the relation between the personality of authors and their motivation for blogging [8]. In addition to pure textual content, links between personality and user activities in social networks have been studied too, revealing for instance a significant relation between extraversion of a user and the number of friends he has, or the number of mobile phone calls he makes [5, 7]. In this paper we focus on multimodal personality impression recognition of video bloggers (vloggers). Analysis of video content appears to be one of the least studied problems in the domain of computational personality recognition [4].

In machine learning approaches to personality recognition from user generated content, ground truth data is usually collected by asking the users who generated the content to take a Big Five inventory survey. A nice example of such ground truth data collection is myPersonality [13]. The work in this paper is different in the sense that the ground truth data does not come from the vloggers themselves, but from other users watching the videos. In other words, the task that we address is not recognition of the true personality traits of vloggers, but *predicting how the personality of vlog-*

gers is perceived by their viewers. To this extent we use both non-verbal cues, i.e. audio-video features, as well as textual analysis of the transcripts of the videos.

Given a video, the aim is to obtain 5 scores based on the 10-item measure of the Big Five¹ (or Five-Factor Model) dimensions. We treat this problem as a multivariate regression task, in which we make a combined prediction for all 5 personality trait scores, instead of training a regressor for each trait separately. Some initial research has been done on the use of multivariate regression for personality prediction on Facebook [1, 10] and Sina Microblog [3]. In the current paper we investigate whether the promising trend of good results can be extended to perceived personality prediction of vloggers. In particular, we measure the performance of 5 multivariate regression techniques [18] on a YouTube personality dataset [4]. We contrast these 5 multivariate regression techniques with a single target approach in which a model is trained for each target variable separately, as well as with a baseline algorithm.

The remainder of this paper is structured as follows: first we describe the dataset and the features in Section 2. In Section 3 we explain the 5 multivariate regression models that we study. The performance of these regression models is discussed in Section 4. In Section 5 we conclude with our overall findings.

2. DATASET AND FEATURES

The YouTube personality dataset² consists of a collection of *audio-video features*, *speech transcripts*, *gender*, and *personality impression scores* for a set of 404 YouTube vloggers. The vloggers explicitly show themselves in front of a webcam, talking about a variety of topics including personal issues, politics, movies, books, etc. The *audio-video features* were automatically extracted from the conversational excerpts of vlogs and aggregated at the video level. The *speech transcripts* correspond to the full video duration. The transcripts are provided in raw text and contain a total of approx. 10K unique words and approx. 240K word tokens. Finally, the *personality impressions* consist of Big Five personality scores that were collected using Amazon Mechanical Turk and the Ten-Item Personality Inventory (TIPI). MTurk annotators watched one-minute slices of each vlog, and rated impressions using a personality questionnaire.

In addition to the audio-video features and the gender of the vlogger, we extracted a wide variety of linguistic and emotional features from the transcripts of the vlogs. Overall, we used the following features³ (including the features provided with the dataset):

1. **Gender:** We used a binary *gender* feature to identify male and female. Overall, the data is balanced in terms of gender distribution and includes 210 females (52%) and 194 males (48%).
2. **Audio-Video:** We used all 25 audio-video features that are provided with the vlog dataset [4]. These include speaking activity and prosody cues such as speaking time and pitch, as well as video features such as the number of look turns and camera proximity.

3. **LIWC:** From the transcripts we extracted **81 features** using the Linguistic Inquiry and Word Count (LIWC) tool [17], including features related to standard counts (e.g., word count), psychological processes (e.g., the number of anger words such as *hate* and *annoyed* in the transcript), relativity (e.g., the number of verbs in the future tense), personal concerns (e.g., the number of words that refer to occupation such as *job* and *majors*), and linguistic dimensions (e.g., the number of swear words).
4. **NRC:** NRC is a lexicon that contains more than 14,000 distinct English words annotated with 8 emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust), and 2 sentiments (negative, positive) [14]. For each transcript we counted the number of words in each of the 8 emotion and sentiment categories, resulting in **10 features** per transcript.
5. **MRC:** MRC is a psycholinguistic database⁴ which contains psychological and distributional information about words. The MRC database contains 150,837 entries with information about 26 properties (e.g., the number of syllables in the word, the number of letters, etc.), although not all properties are available for every word. Using MRC we generated **14 features** for every transcript (see Table 11 in the appendix) by adding the MRC-scores for each word in the transcript.
6. **SentiStrength:** SentiStrength⁵ assigns to each text a positive, negative and neutral sentiment score on a scale of 1 (no sentiment) to 5 (very strong sentiment). Texts may be simultaneously positive, negative and neutral. We used SentiStrength to compute 3 sentiment scores (**3 features**) for every transcript.
7. **SPLICE:** We used SPLICE⁶ (Structured Programming for Linguistic Cue Extraction) to extract **74 linguistic features**, including cues that relate to the positive or negative self evaluation of the speaker (e.g., *I'm able*, *don't know*), complexity and readability scores.

The underlying rationale for including linguistic and emotional features is that people with different personality traits will express themselves differently and, hence, will use different words (phrases) and emotions (anger, joy) when expressing themselves. A relation between emotions and personality traits has been observed in past research as well [6]. We used Spearman's rank correlation coefficient to assess the strength of the relationship between the different features described above and the 5 perceived personality traits (see the appendix). Given the highly skewed distribution of some of the features, we decided to report all the correlations using Spearman's coefficient, which is better suited for non-normal data. Our results indicate a strong relationship between many linguistic and emotional features and personality impressions. Motivated by previous research, and the observed correlation between features and personality impressions, we decided to include these features in our regression models. Some of our feature sets have some semantic overlap (e.g. NRC and SentiStrength, and LIWC and

¹http://www.sjdm.org/dmidi/Ten_Item_Personality_Inventory.html

²<https://www.idiap.ch/dataset/youtube-personality>

³See the appendix for a full list of features and how they correlate with the 5 perceived personality traits. All features except for the gender feature are numerical.

⁴http://www.psych.rl.ac.uk/User_Manual_v1_0.html

⁵<http://sentistrength.wlv.ac.uk>

⁶<http://splice.cmi.arizona.edu>

Table 1: Pearson product-moment correlation results among personality impression scores on 5 traits: *Extraversion (Extr)*, *Agreeableness (Agr)*, *Conscientiousness (Cons)*, *Emotional Stability (EmoStab)*, *Openness (Open)*. Significant correlations ($p < .05$) among the personality impressions are indicated in bold.

	Extr	Agr	Cons	EmoStab	Open
Extr	1.00				
Agr	.02	1.00			
Cons	-.03	.38	1.00		
EmoStab	.06	.69	.54	1.00	
Open	.56	.29	.26	.30	1.00

SPLICE). The use of feature selection methods to be more selective in the choice of features is an interesting direction for further research.

3. MULTIVARIATE REGRESSION

As the results in Table 1 indicate, there is a clear correlation among different personality trait impression scores in the YouTube dataset. This makes personality score prediction a good candidate for multivariate regression, where the dependencies between the target variables are taken into account to make a combined prediction. Formally, multivariate regression addresses the following problem:

Let F be the vector (feature space or input space) including m features, f_1, f_2, \dots, f_m , and T be the target vector (output space) including n target variables t_1, t_2, \dots, t_n . The goal of a multivariate regression algorithm is to learn a model $M : F \rightarrow T$ that minimizes the prediction error over a test test. Using this formulation, the 6 multivariate regression algorithms that we use in this paper are [18]:

1. **Single Target (ST)**: In ST, for each target variable t_i , a single model is trained based on the input vector F (i.e., $F \rightarrow t_i$). The results of the multi-target model are comprised of all n single target ones.
2. **Multi-Target Stacking (MTS)**: MTS consists of two steps. In the first step, n single-target models are used as in ST, however, MTS includes an additional step where the input space for each target variable is expanded by the predicted results of the other target variables ($n - 1$ predicted values) from step one. Let t'_1, t'_2, \dots, t'_n be the prediction results from the first step, then for example the input space for t_1 in step two is $[f_1, f_2, \dots, f_m, t_2, t_3, \dots, t_n]$.
3. **Multi-Target Stacking Corrected (MTSC)**: In MTSC, an internal cross validation sampling technique is used to avoid over-estimation of the training set. In MTSC, by using k -fold sampling, the prediction results of $\frac{k-1}{k}$ % of the whole training set are used to expand the input space in the second step as in MTS.
4. **Ensemble of Regressor Chains (ERC)**: The idea behind ERC is chaining single-target regression models. By choosing an order for the target variables (e.g., $O = (t_1, t_2, \dots, t_n)$), the learning model for each target variable t_j relies on the prediction results of all target variables t_i which appear before t_j in the list. For the first target variable, a single-target regression model as in ST predicts the value, then the input space for

the next target variable is extended with the prediction results of the previous one and so on. Since in this model the order of the chosen chain affects the results, the average prediction result of r different chains (typically $r = 10$) for each target variable is used as the final prediction result.

5. Ensemble of Regressor Chains Corrected

(**ERCC**): The difference between ERC and ERCC is similar to that between MTS and MTSC, i.e. the use of k -fold sampling to increase the reliability of the predictions based on the training set.

6. Multi-objective random forest (MORF):

MORF is based on ensembles of multi-objective decision trees. We refer to [12] for further explanation.

Note that ST does not leverage the prediction result for one personality trait to make a prediction for another, while all other algorithms (MTS, MTSC, ERC, ERCC and MORF) do in one way or another. For the results in Section 4 we used the implementation of these algorithms in Mulan.⁷ All algorithms except MORF use Weka decision trees as a base learner. For further information we refer to [18].

4. PERFORMANCE

We measured the performance of the prediction algorithms using the root mean square error (RMSE) and the coefficient of determination (R^2). To this end, we used the split of the data into 348 training and 56 test instances that was suggested by the organizers of WCP2014. In addition to results for the 6 algorithms described in Section 3, in Table 2 we also include results for a baseline algorithm that predicts for each personality trait score the average of the scores in the training set for that trait. To measure significant differences in prediction errors between the learned models and the baseline, we conducted two-tailed paired t-tests for the RMSE, and two-tailed single t-tests for R^2 . All results are obtained with Mulan⁸ and compared against the mean value baseline.

It can be seen from the results in Table 2 that all 6 algorithms outperform (i.e., have a lower prediction error than) the baseline model for all 5 personality types. In addition, positive values for R^2 are also observed for all the algorithms which further indicates better performance than the average baseline model ($0\% \leq R^2 \leq 37\%$). For *Extraversion*, ST, MTSC and ERCC algorithms performed best ($R^2 = 20\%$), while for *Agreeableness*, ST and ERCC were the best performers ($R^2 = 37\%$). In the case of *conscientiousness*, MTSC and MORF emerged as the best models with lowest prediction errors ($R^2 = 16\%$). However, MTS and ERC showed performance with significant difference ($14\% \leq R^2 \leq 16\%$). Interestingly, MORF outperformed for this personality type, but performed poorly for the others. In particular, it failed to fit the data correctly for Emotional Stability and Openness ($R^2 = 0\%$). *Emotional Stability* had three main winners – ST, MTSC and ERCC ($R^2 = 13\%$). Finally, in case of *openness* personality type, MTS had the lowest prediction error when compared to the baseline ($R^2 = 14\%$). The difference in the error between MTS and the baseline was also found to be significant. In

⁷<http://mulan.sourceforge.net/>

⁸Except significance tests results, which were obtained in R.

Table 2: Root mean square error (RMSE) and Coefficient of determination (R^2) results for personality trait prediction using multivariate regression algorithms. In each column, the lowest error and highest determination are typeset in bold. Significant differences ($p < .05$) are marked using *.

Feature Set	Extr		Agr		Con		EmoStab		Open	
	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2
Average Baseline	1.02		.91		.71		.75		.83	
ST	.91	20	.72	37	.65	16	.70	13	.80	7
MTS	.95	13	.76	30	.66*	14*	.71	10	.77*	14*
MTSC	.91	20	.73	35	.64	18	.70	13	.79	9
ERC	.93	17	.74	34	.65*	16*	.72	8	.79*	9*
ERCC	.91	20	.72	37	.65	16	.70	13	.80	7
MORF	.98	7	.84	15	.64	18	.75	0	.83	0

addition to MTS, ERC also performed better than the baseline with significant difference.

Another interesting observation is that, in past studies, models for Extraversion personality trait often performed best, while models for Agreeableness performed worst (e.g. [2], [4]). However, our results show that models for Agreeableness were the best performers while models for Extraversion came second. The prime difference in the settings of the previous studies (mentioned above) and ours are: the set of features used (in both cases), the dataset (in [2]), among others. Therefore, it is hard to put a finger on the exact cause for this change in the performance. This could be an interesting problem to investigate in future. Finally, our overall prediction results are important because previously published methods [4] for the same dataset show an improvement over the baseline for the majority of personality traits, but not for “all”. Furthermore, based on our results we can say that multivariate regression does show potential for solving personality prediction problems. However, it is possible for single target methods to give equally good results.

5. CONCLUSION

In this study, we explored the use of multivariate regression algorithms to predict how the personality of YouTube vloggers is perceived by their viewers. Instead of training 5 learners separately to predict the 5 personality scores, multivariate regression techniques make a combined prediction. Given the correlation among different personality traits, this sounds promising. However, although the multivariate regression techniques that we evaluated performed well on a YouTube personality dataset of 404 vlogs, they did not clearly outperform a single target approach in which a model was trained for each personality trait separately. All 6 models outperformed our average prediction baseline though for all 5 personality dimensions. Past methods proposed for the same dataset were able to outperform the baseline for 3 personality traits simultaneously only.

Overall we used a wide variety of features common to all approaches that we evaluated, including gender, audio-video features, linguistic features and emotional features. An interesting direction for future research would be to investigate whether the application of feature selection techniques can further improve the results.

6. REFERENCES

- [1] Y. Bachrach, M. Kosinski, T. Graepel, P. Kohli, and D. Stillwell. Personality and patterns of Facebook usage. In *Proc. of ACM Web Science*, pages 24–32, 2012.
- [2] Y. Bachrach, M. Kosinski, T. Graepel, P. Kohli, and D. Stillwell. Personality and patterns of facebook usage. In *Proceedings of the 4th Annual ACM Web Science Conference*, pages 24–32. ACM, 2012.
- [3] S. Bai, B. Hao, A. Li, S. Yuan, R. Gao, and T. Zhu. Predicting Big Five personality traits of microblog users. In *Proc. of IEEE/WIC/ACM WI-IAT*, volume 1, pages 501–508, 2013.
- [4] J.-I. Biel and D. Gatica-Perez. The YouTube lens: Crowdsourced personality impressions and audiovisual analysis of vlogs. *IEEE Transactions on Multimedia*, 15(1):41–55, 2013.
- [5] G. Chittaranjan, J. Blom, and D. Gatica-Perez. Mining large-scale smartphone data for personality studies. *Personal and Ubiquitous Computing*, 17(3):433–450, 2013.
- [6] G. Farnadi, G. Sitaraman, M. Rohani, M. Kosinski, D. Stillwell, M. Moens, S. Davalos, and M. De Cock. How are you doing? Emotions and personality in Facebook. In *Proc. of EMPIRE*, pages 45–56, 2014.
- [7] G. Farnadi, S. Zoghbi, M. Moens, and M. De Cock. Recognising personality traits using Facebook status updates. In *Proc. of WCPR*, pages 14–18, 2013.
- [8] A. J. Gill, S. Nowson, and J. Oberlander. What are they blogging about? Personality, topic and motivation in blogs. In *Proc. of ICWSM*, 2009.
- [9] R. Hu and P. Pu. Enhancing collaborative filtering systems with personality information. In *Proc. of ACM RecSys*, pages 197–204, 2011.
- [10] F. Iacobelli and A. Culotta. Too Neurotic, Not Too Friendly: Structured Personality Classification on Textual Data. In *Proc of Workshop on Computational Personality Recognition, AAAI Press, Melon Park, CA*, pages 19–22, 2013.
- [11] O. P. John and S. Srivastava. The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of Personality: Theory and Research*, 2:102–138, 1999.
- [12] D. Koccev, C. Vens, J. Struyf, and S. Džeroski. Ensembles of multi-objective decision trees. In *Proc. of ECML*, pages 624–631, 2007.
- [13] M. Kosinski, D. Stillwell, and T. Graepel. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the PNAS*, 110(15):5802–5805, 2013.
- [14] S. Mohammad, X. Zhu, and J. Martin. Semantic role labeling of emotions in tweets. In *Proc. of WASSA*, pages 32–41, 2014.
- [15] J. Oberlander and S. Nowson. Whose thumb is it anyway?: Classifying author personality from weblog text. In *Proc. of COLING/ACL*, pages 627–634, 2006.
- [16] R. D. Oliveira, M. Cherubini, and N. Oliver. Influence of personality on satisfaction with mobile phone services. *TOCHI*, 20(2), 2013.
- [17] J. Pennebaker and M. Francis. *Linguistic Inquiry and Word Count*. Taylor & Francis, 1999.
- [18] E. S. Xioufis, W. Groves, G. Tsoumakas, and I. P. Vlahavas. Multi-label classification methods for multi-target regression. *CoRR*, abs/1211.6581, 2012.

APPENDIX

The Spearman's rank correlation coefficients of all the extracted features with the 5 personality impression scores (*Extraversion (E)*, *Agreeableness (A)*, *Conscientiousness (C)*, *Emotional Stability (ES)*, *Openness (O)*) are presented in Table 3–11. Significant ($p < 0.05$) correlations between features and personality impression scores are typeset in bold.

Table 3: Correlation results between the SPLICE features and the 5 personality impression scores. 1-35 Features

Splice Features	E	A	C	ES	O
numChars	0.08	-0.10	0.12	0.00	-0.03
numCharsMinusSpaces					
AndPunctuation	0.08	-0.10	0.13	0.01	-0.03
numWords	0.08	-0.10	0.11	0.00	-0.03
numSentences	0.13	-0.14	0.00	-0.05	-0.04
numPunctuation	0.10	-0.12	-0.01	-0.03	-0.04
numNouns	0.10	-0.09	0.09	0.00	-0.03
nounRatio	0.13	0.03	-0.16	0.00	0.02
numVerbs	0.07	-0.09	0.10	-0.01	-0.03
verbRatio	-0.04	-0.01	-0.04	-0.08	-0.02
numAdjectives	0.10	-0.11	0.10	-0.01	-0.01
adjectiveRatio	0.10	-0.05	0.00	0.00	.07
numAdverbs	0.05	-0.08	0.09	-0.01	-.03
adverbRatio	-0.06	0.03	-0.05	-0.04	0.00
firstPersonSingular	0.06	-0.01	-0.04	-0.03	-.01
firstPersonPlural	0.07	-0.05	0.14	0.04	-0.01
secondPerson	0.14	-0.04	0.10	0.02	0.00
thirdPersonSingular	-0.02	-0.18	0.10	-0.01	-0.05
thirdPersonPlural	-0.02	-0.18	0.10	-0.01	-0.05
iCanDoIt	0.01	-0.06	-0.01	-0.01	-0.05
doKnow	0.06	0.01	-0.03	-0.01	-0.07
posSelfImage	0.07	0.03	-0.03	0.02	0.02
iCantDoIt	-0.02	-0.06	-0.07	-0.04	-0.07
dontKnow	0.03	-0.02	-0.09	-0.02	0.04
negSelfImage	0.01	-0.11	-0.09	-0.07	0.01
numImperatives	0.16	-0.03	-0.03	0.00	-0.01
suggestionPhrases	-0.02	-0.02	0.14	0.05	-0.06
inflexibility	0.01	-0.02	0.02	-0.06	-0.10
contradict	0.03	-0.05	0.01	0.02	0.00
totalDominance	0.11	-0.04	0.04	0.01	-0.06
dominanceRatio	0.03	0.10	0.06	0.10	-0.03
numAgreement	0.09	-0.07	-0.05	0.01	-0.03
agreementRatio	0.05	-0.01	-0.16	0.02	-0.01
totalSubmissiveness	0.04	-0.10	-0.14	-0.07	-0.02
submissivenessRatio	-0.01	-0.04	-0.14	-0.04	-0.01

Table 4: Correlation results between gender and the 5 personality impression scores. The absolute value of Spearman's rank c.c. is reported here.

Feature	E	A	C	ES	O
gender	.02	0.21	.02	.02	.07

Table 5: Correlation results between the SPLICE features and the 5 personality impression scores. 36 - 74 Features.

Splice Features	E	A	C	ES	O
Imagery	0.19	-0.02	0.07	0.01	0.11
Pleasantness	0.13	0.26	0.05	0.20	0.17
Activation	0.11	-0.08	-0.01	-0.09	0.06
avgWordLength	0.02	-0.05	0.32	0.12	0.03
avgSentenceLength	-0.14	0.09	0.23	0.14	-0.03
numSyllables	0.08	-0.10	0.12	0.01	-0.03
avgSyllablesPerWord	-0.01	0.03	0.28	0.14	-0.01
numWords3OrMoreSyll	0.07	-0.11	0.18	0.03	-0.02
rateWords3OrMoreSyll	0.00	-0.03	0.23	0.12	0.02
numWords6OrMoreChars	0.06	-0.11	0.15	0.01	-0.03
rateWords6OrMoreChars	-0.04	-0.04	0.22	0.07	0.01
numWords7OrMoreChars	0.07	-0.10	0.18	0.02	-0.02
rateWords7OrMoreChars	0.01	0.01	0.29	0.12	0.06
LexicalDiversity	-0.05	0.06	-0.06	0.01	0.05
hedgeVerb	-0.01	-0.03	0.10	0.04	-0.06
hedgeConj	-0.02	-0.07	0.06	0.01	-0.04
hedgeAdj	-0.05	-0.10	0.08	0.00	-0.08
hedgeModal	-0.05	-0.03	0.16	0.04	0.00
hedgeAll	-0.04	-0.10	0.11	0.01	-0.04
numDisfluencies	-0.01	0.01	-0.05	0.06	-0.06
disfluencyRatio	-0.07	0.10	-0.12	0.08	-0.07
numInterjections	0.11	-0.11	-0.11	-0.09	-0.03
interjectionRatio	0.06	-0.08	-0.28	-0.13	-0.03
numSpeculate	-0.03	-0.16	0.02	0.00	-0.08
Expressivity	-0.11	0.09	-0.09	0.00	0.03
Pausality	0.13	-0.07	0.02	-0.06	0.06
questionCount	0.22	-0.22	-0.11	-0.14	-0.01
questionRatio	0.20	-0.21	-0.15	-0.15	0.01
pastTense	0.02	-0.03	0.08	-0.01	-0.05
presentTense	0.09	-0.11	0.07	-0.03	-0.03
ARI	-0.10	0.06	0.30	0.14	-0.01
FRE	0.09	-0.07	-0.31	-0.16	0.02
CLI	0.14	-0.09	-0.23	-0.14	0.03
LWRF	-0.13	0.09	0.25	0.15	-0.02
FOG	-0.11	0.07	0.28	0.16	-0.01
SMOG	-0.07	0.04	0.29	0.16	0.00
DALE	-0.14	0.09	0.23	0.14	-0.02
LIX	-0.12	0.06	0.29	0.13	0.00
RIX	-0.10	0.08	0.32	0.17	0.01
FRY	-0.13	0.09	0.26	0.15	-0.02

Table 6: Correlation results between the audio-video features and the 5 personality impression scores.

Speaking Activity	E	A	C	ES	O
Time speaking	0.18	0.07	0.27	0.15	0.14
Avg length segm	0.17	0.04	0.17	0.11	0.12
No. of turns	-0.17	0.01	-0.04	-0.04	-0.09
Prosodic Cues	E	A	C	ES	O
mean pitch	0.23	0.09	-0.08	-0.06	0.07
sd pitch	-0.12	-0.17	0.01	-0.02	-0.02
meanconf pitch	0.21	0.14	0.00	0.01	0.04
sdconf pitch	0.16	0.11	0.03	0.02	0.06
mean spec entropy	0.11	-0.05	-0.10	0.00	-0.02
sd spec entropy	0.02	0.01	0.02	-0.05	0.02
mean val apeak	-0.03	0.10	0.01	0.01	-0.04
sd val apeak	0.00	-0.06	0.01	0.01	0.04
mean loc apeak	0.29	0.02	-0.06	-0.08	0.05
sd loc apeak	-0.03	-0.13	-0.10	-0.08	-0.06
mean num apeak	0.17	-0.05	-0.08	-0.06	-0.02
sd num apeak	0.06	-0.10	-0.10	-0.07	-0.08
mean energy	0.24	-0.10	-0.08	-0.03	0.04
sd energy	0.10	-0.04	-0.09	-0.04	0.09
mean d energy	-0.06	0.02	0.03	0.02	-0.06
sd d energy	0.31	-0.12	-0.11	-0.05	0.08
avg voiced seg	-0.05	-0.12	-0.07	-0.06	-0.08
voice rate	0.01	0.11	0.04	0.04	0.05
Visual Activity	E	A	C	ES	O
hogv entropy	0.32	-0.04	-0.22	-0.08	0.19
hogv median	0.29	0.03	-0.16	-0.02	0.22
hogv cogR	0.01	-0.06	-0.02	0.05	0.01
hogv cogC	0.00	-0.01	-0.04	-0.06	-0.06

Table 7: Correlation results between the LIWC features and the 5 personality impression scores. 1-25 Features.

LIWC Features	E	A	C	ES	O
Word Count	-0.07	0.10	-0.15	0.01	0.05
Words Per Sentence	0.13	-0.06	-0.26	-0.11	0.01
Sixltr	-0.01	0.05	-0.27	-0.07	-0.03
Dictionary words	0.22	-0.08	0.01	0.02	0.10
Numerals	0.02	-0.05	-0.03	-0.07	0.10
funct	0.18	-0.07	-0.08	-0.03	0.07
Pronouns	0.04	-0.08	0.15	0.03	0.07
ppron	-0.05	-0.08	0.18	0.08	0.02
i	0.03	-0.12	0.23	0.06	0.03
we	-0.07	0.03	-0.10	-0.04	-0.02
you	-0.14	-0.02	-0.02	-0.02	-0.07
shehe	-0.09	0.10	0.01	0.07	-0.03
they	0.06	0.14	-0.11	0.03	0.03
ipron	0.15	0.00	-0.03	-0.08	0.04
article	0.00	0.08	-0.12	-0.09	-0.02
verb	0.08	-0.02	0.19	0.07	0.08
auxverb	0.06	0.00	0.18	0.08	0.06
past	0.08	-0.02	0.06	0.04	0.05
present	-0.02	0.01	0.17	0.07	0.04
future	0.09	-0.02	0.01	-0.04	0.08
adverb	0.05	-0.11	0.13	-0.03	0.06
preps	0.06	0.01	-0.28	-0.07	-0.08
conj	0.08	-0.16	-0.05	-0.02	0.04
negate	0.03	0.17	0.22	0.18	0.05
quant	0.09	0.15	-0.10	0.02	0.06

Table 8: Correlation results between the LIWC features and the 5 personality impression scores. 26-51 Features.

LIWC Features	E	A	C	ES	O
number	-0.05	0.03	-0.06	-0.05	0.03
swear	-0.08	0.39	0.20	0.25	0.04
social	-0.07	0.05	-0.16	0.00	-0.03
family	-0.07	0.02	-0.07	0.05	-0.04
friend	0.00	-0.09	-0.01	0.00	-0.12
humans	-0.12	0.08	-0.02	0.02	-0.06
affect	-0.12	0.00	0.06	0.06	-0.04
posemo	-0.09	-0.27	-0.08	-0.15	-0.12
negemo	-0.03	0.40	0.27	0.35	0.14
anx	-0.06	0.04	-0.01	0.07	-0.02
anger	-0.07	0.41	0.27	0.27	0.07
sad	0.06	0.00	0.03	0.03	0.13
cogmech	0.20	-0.09	-0.09	-0.02	0.08
insight	0.04	0.02	-0.06	-0.02	0.01
cause	0.08	-0.02	0.07	0.03	0.05
discrep	0.12	-0.08	-0.11	-0.11	-0.02
tentat	0.21	0.05	-0.02	-0.08	0.07
certain	-0.01	0.07	-0.04	0.07	0.02
inhib	0.09	0.09	-0.01	0.08	0.07
incl	-0.04	-0.14	-0.23	-0.01	-0.03
excl	0.14	-0.06	0.08	0.00	0.10
percept	-0.06	-0.02	0.02	-0.05	-0.06
see	-0.08	0.01	0.01	-0.03	-0.03
hear	-0.01	0.09	0.00	0.00	-0.08
feel	0.05	-0.09	0.03	0.03	0.05
bio	-0.05	0.12	0.12	0.19	0.08

Table 9: Correlation results between the LIWC features and the 5 personality impression scores. 52-81 Features.

LIWC Features	E	A	C	ES	O
body	0.00	0.15	0.18	0.15	0.08
health	0.03	0.08	-0.02	0.11	0.10
sexual	-0.19	0.17	0.13	0.22	-0.02
ingest	-0.04	-0.04	-0.07	-0.04	-0.03
relativ	-0.08	0.00	-0.01	0.07	-0.05
motion	-0.06	0.02	0.03	0.10	-0.11
space	-0.15	0.03	-0.10	-0.01	-0.08
time	0.05	-0.03	0.15	0.08	0.07
work	0.08	-0.13	-0.24	-0.13	-0.05
achieve	0.02	-0.06	-0.20	-0.10	0.07
leisure	-0.06	-0.06	0.01	-0.12	-0.14
home	-0.02	-0.01	-0.02	0.03	0.04
money	-0.05	-0.02	-0.10	-0.07	-0.02
relig	-0.13	0.11	0.05	0.03	-0.02
death	0.04	0.16	0.00	0.10	0.06
assent	-0.17	-0.03	0.26	0.05	-0.08
nonfl	0.18	-0.13	0.06	-0.09	0.14
filler	-0.03	-0.03	0.23	0.07	0.05
Period	-0.05	0.01	0.22	0.08	0.00
Comma	0.03	-0.01	0.14	-0.02	0.04
Colon	0.00	0.04	-0.08	0.01	0.10
SemiC	-0.06	0.07	0.05	0.03	0.03
QMark	-0.22	0.19	0.19	0.15	-0.03
Exclam	-0.21	-0.01	0.08	0.00	-0.13
Quote	0.01	0.02	-0.01	0.04	-0.02
Dash	0.10	0.10	0.15	0.07	0.11
Apostro	0.00	0.02	0.26	0.08	0.07
Parenth	-0.02	0.03	-0.01	0.08	0.00
OtherP	-0.04	-0.06	-0.07	-0.10	-0.09
AllPct	0.00	0.07	0.31	0.08	0.06

Table 10: Correlation results between the emotional features and the 5 personality impression scores.

NRC	E	A	C	ES	O
Positive	0.07	-0.01	0.16	0.08	0.02
Negative	0.07	-0.29	-0.05	-0.19	-0.11
Anger	0.10	-0.29	-0.06	-0.16	-0.05
Anticipation	0.02	-0.01	0.13	0.05	-0.06
Disgust	0.03	-0.29	-0.05	-0.18	-0.10
Fear	0.03	-0.20	0.03	-0.14	-0.08
Joy	0.09	0.07	0.15	0.10	0.03
Sadness	0.03	-0.21	0.03	-0.14	-0.06
Surprise	0.06	0.03	0.12	0.08	-0.05
Trust	0.03	-0.05	0.09	0.04	-0.02
SentiStrength	E	A	C	ES	O
Positive	0.04	0.35	0.15	0.27	0.10
Neutral	0.00	-0.34	-0.19	-0.35	-0.20
Negative	-0.03	-0.12	-0.05	-0.04	0.01

Table 11: Correlation results between the MRC features and the 5 personality impression scores.

MRC	E	A	C	ES	O
NLET	-0.04	-0.09	0.30	0.07	0.00
NPHON	-0.11	-0.04	0.17	0.08	-0.05
NSYL	-0.14	-0.02	0.16	0.04	-0.07
KF FREQ	-0.03	-0.01	0.25	0.08	-0.02
KF NCATS	-0.13	0.02	0.07	-0.02	-0.02
KF NSAMP	-0.17	0.04	0.15	0.02	-0.05
TL FREQ	-0.03	0.01	0.28	0.09	-0.01
BROWN FREQ	-0.09	0.10	-0.06	-0.02	-0.07
FAM	-0.05	0.05	0.08	0.01	0.00
CONC	0.05	0.01	-0.06	-0.04	0.02
IMAG	0.08	0.03	-0.04	-0.04	0.03
MEANC	0.01	0.04	-0.11	-0.08	-0.01
MEANP	0.15	0.05	0.10	0.05	0.07
AOA	0.06	-0.07	0.18	0.09	0.00