Bottom-Up and Top-Down: Predicting Personality with Psycholinguistic and Language Model Features

Yash Mehta*
Gatsby Computational Neuroscience Unit
University College London

Samin Fatehi*
School of Computer Engineering
Iran University of Science and Technology

Amirmohammad Kazameini
School of Computer Engineering
Iran University of Science and Technology

Clemens Stachl
Department of Communication
Stanford University

Erik Cambria
School of Computer Science and Engineering
Nanyang Technological University

Sauleh Eetemadi
School of Computer Engineering
Iran University of Science and Technology

***Abstract*—State-of-the-art personality prediction with text data mostly relies on bottom up, automated feature generation as part of the deep learning process. More traditional models rely on hand-crafted, theory-based text-feature categories. We propose a novel deep learning-based model which integrates traditional psycholinguistic features with language model embeddings to predict personality from the Essays dataset for Big-Five and Kaggle dataset for MBTI. With this approach we achieve state-of-the-art model performance. Additionally, we use interpretable machine learning to visualize and quantify the impact of various language features in the respective personality prediction models. We conclude with a discussion on the potential this work has for computational modeling and psychological science alike.[1]**

***Index Terms*—Language Models, Automated Personality Prediction, Psycholinguistic Features, NLP**

I. Introduction

Personality traits are generally referred to as relatively stable patterns of thoughts, feelings, and behaviors that have been associated with a wide range of important life outcomes and choices [1], [2]. Specifically, personality traits have repeatedly been related to individual (e.g., well-being, psychopathology), inter-personal (e.g., relationship satisfaction), and social-institutional outcomes (e.g., occupational choices, job success; [3], [4]). Hence, there is an increasing interest to develop models that can use online data on human behavior and preferences (i.e., digital footprints) to automatically predict individuals' levels of personality traits for use in recommender-systems [5], [6], product and service personalization [7], [8] job screenings [9], social network analysis [10], and sentiment analysis [11].

A. Personality Theories and Assessment

Across time, numerous taxonomies and models for the comprehensive and systematic description of human personality have been proposed [12]. The five factor model (Big Five) is most widely accepted in psychological science [13] and consists of five broad dimensions of personality (Openness to Experience, Conscientiousness, Agreeableness, Extraversion, and Neuroticism or positively keyed, emotional stability).

*These authors contributed equally

[1]Code: https://github.com/yashsmehta/personality-prediction

To get an estimate of an individual's scores on each of these dimensions, standardized self-report questionnaires are used (e.g., NEO-PI-3, [14]). While personality assessment based on the five factor, trait model is most commonly used in personality science, the Myers–Briggs type indicator is another widely used questionnaire in applied settings [15]. Unlike the Big Five personality trait taxonomy (which conceptualizes personality as latent trait scores), MBTI describes personality in the form of 16 types that are created from the combination of binary assignments to four dimensions: introversion versus extraversion, sensing versus intuiting, thinking versus feeling, and judging versus perceiving [15]. The Myers–Briggs Type Indicator (MBTI) has been heavily criticized due to a multitude of methodological shortcomings [16]. Still, it remains one of the most widely administered personality inventories in the world year [17].

B. Ethical Considerations

In classical personality assessment, self-report questionnaires are used to get an estimate for people's assumed latent trait levels. However, recent developments in the area of automated personality prediction, suggest that digital footprints and behavioral data can be used to automatically infer peoples self-reported personality trait levels with some degree of accuracy and without explicit consent [18]–[20]. Computational personality assessment is appealing, because it holds the promise to remove the need to fill in questionnaires. While the performance of these models is not high enough to allow for the precise distinction of people based on their traits, predictions can still be "right" on average and be utilized for personalizing services and products and for digital mass persuasion [6]. In that regard, computational personality trait assessment also raises serious concerns with regard to individual privacy and the conception of informed consent [21].

C. Personality and Language Use

Individual differences in language use have been considered as reflections of psychological phenomena since the early days of Freud [22]. In the last decade, numerous empirical studies have linked peoples' language use to their self-reported personality trait scores [20], [23]–[25]. For example, individuals scoring higher on extraversion were found to use more positive emotion words (e.g., great, amazing, happy) whereas those higher in neuroticism were found to use first-person singulars (e.g., I, mine, me) more frequently [26]. Initially, these findings led to the development of psycho-linguistic word-categorizations (e.g., Linguistic Inquiry and Word Count - LIWC [26], [27]) to allow for the systematic analysis of language data in psychology.

D. Rationale

One of the latest and most promising developments in language-based personality assessment is the use of transfer learning techniques. A language model is pre-trained using unsupervised learning on large amounts of unlabeled data to gain an understanding of the underlying structure of the language. These language models have been used to obtain state-of-the-art results across many famous NLP benchmarks including GLUE [28] and SQuAD [29]. In this paper, we leverage the power of these language models, perform extensive empirical experiments and achieve state-of-the-art results across the famous Essays [26] and Kaggle[2] personality datasets. We also study the contributions of traditional hand engineered psycholinguistic features by analyzing the effects of individual psycholinguistic features on predicting a particular personality trait. Additional resources, syntax, and data are available in our open-science repository for transparency and full reproducibility: https://osf.io/rg5tf/.

II. Related Work

A large amount of research has been dedicated for automated personality prediction from the text modality. Earlier works on author personality prediction focused on extracting features from text based on the lexicon, syntax, writing style and topic, followed by seeing which of these features are highly correlated with personality traits using a correlation metric such as the Pearson correlation [30]. Empirical results using LIWC demonstrate its ability to detect meaning in a wide variety of experimental settings, including to show attentional focus, emotionality, social relationships, thinking styles, and individual differences. Mairesse et al. [31] developed a document-level feature set for personality prediction, consisting of 84 features. These text features are then fed into traditional machine learning classifiers such as logistic regression, support vector machine (SVM) [32], Naïve Bayes, etc for getting the final personality prediction.

More recent work rely on the advances in deep learning and make use of pre-trained word embeddings like Word2Vec [33] and Glove [34] to build better performing personality prediction models. It was found that combining commonsense knowledge with psycho-linguistic features resulted in a remarkable improvement in the accuracy [35]. Another work in this direction is the famous 1-D CNN n-grams model proposed by Majumder et al. [36], which achieved the state-of-the-art personality prediction performance, until beaten by the language model based ensemble method (BB-SVM) by Kazemeini et al. [37]. Recently, Mehta et al. [38] reviewed the latest advances in deep learning-based automated personality with a focus on effective multimodal personality prediction.

[2]https://www.kaggle.com/datasnaek/mbti-type

III. Method

A. Datasets

We used the following publicly available personality datasets in our analyses:

1) Essays: The famous stream-of-consciousness dataset consisting of 2468 essays written by students and annotated with the binary labels of the Big Five personality traits which were found by a standardized self-report questionnaire [26].

2) Kaggle MBTI: This data was collected through the PersonalityCafe forum and hence, provides a diverse selection of people interacting in an informal online social setting. This dataset contains 8675 records of the last 50 things an individual posted on the website along with their MBTI binary personality type.

TABLE I
The point-biserial correlation level between the SenticNet values of the document concepts and the Big Five personality traits for the Essays dataset. *p <.05. **p <.001, two-tailed. Correlation coefficients were computed on the complete dataset.

	O	C	E	A	N
Pleasantness	0.041*	0.066*	0.032	0.025	−0.075**
Attention	0.113**	−0.026	0.013	−0.007	−0.017
Sensitivity	−0.011	−0.052*	−0.064*	−0.034	−0.022
Aptitude	−0.045*	0.112**	0.052*	0.081**	−0.020
Polarity	0.000	0.081**	0.037	0.056*	−0.058*

B. Feature Extraction

From the text data, we extract two different types of features, namely, psycholinguistic features (a fixed set of features previously found to have correlations with personality) and language model embeddings.

1) Psycholinguistic Features: We extracted literature-derived psycholinguistic features from the aforementioned text datasets (P = 123). Additionally we retrieved meta-information (called 'readability') from the text and study the degree to which these features are correlated with personality.

- **Mairesse** [31]: A total of 84 features which are made up of LIWC, Medical Research Council [39], prosodic and utterance-type features. These are the widely used 'hand engineered' features in traditional machine learning-based personality prediction models.
- **SenticNet** [40]: A lexicon of over 100,000 commonsense concepts annotated with learnt values of pleasantness, attention, sensitivity, aptitude and polarity. We created our own efficient concept parser to extract these values for the longest length concept. The final value of this sub-feature is the mean of all concepts extracted from the document. The correlation between the these SenticNet features and the Big Five personality traits is shown in Table I.

TABLE II

MODEL	Essays						Kaggle MBTI				
	O	C	E	A	N	Average	I/E	N/S	T/F	P/J	Average
Majority Baseline	51.5	50.8	51.7	53.1	50.0	51.4	77.0	85.3	54.1	60.4	69.2
Majumder et al CNN model [36]	61.1	56.7	58.1	56.7	57.3	58.0	-	-	-	-	-
SOTA [37] [43]	62.1	57.8	59.3	56.5	59.4	59.0	**79.0**	86.0	74.2	65.4	76.1
Psycholinguistic + MLP	60.4	57.3	56.9	57.0	59.8	58.3	77.6	86.3	72.0	61.9	74.5
BERT-base + SVM	63.2	56.2	57.8	57.4	58.8	58.7	77.0	86.2	73.7	60.5	74.4
BERT-base + MLP	**64.6**	**59.2**	**60.0**	**58.8**	60.5	**60.6**	78.3	86.4	74.4	64.4	75.9
All features (base) + MLP	61.1	57.4	57.9	58.6	**60.5**	59.1	78.4	**86.6**	75.9	64.4	76.3
BERT-large + MLP	63.4	58.9	59.2	58.3	58.9	59.7	78.8	86.3	**76.1**	**67.2**	**77.1**

- **NRC Emotion Lexicon** [41]: A lexicon of over 14,000 English words annotated with values of 8 emotions: anger, anticipation, disgust, fear, joy, sadness, surprise and trust. The final value of this sub-feature is a 8 dimension vector, which is the mean of all values of emotionally charged words present in the document.
- **VAD Lexicon** [42]: A lexicon of over 20,000 English words annotated with their valence, arousal and dominance scores. As above, the document VAD value is the mean of all constituent words.
- **Readability**: A number of calculated readability measures based on simple surface characteristics of the text. These measures are basically linear regressions based on the number of words, syllables, and sentences.

2) Language Model Features: We experiment with multiple different language models (BERT [44], Albert [45] and Roberta [46]), but we see similar performance across them and hence only report results for BERT-base and BERT-large in the paper. We perform extensive experiments to arrive at the optimal configuration for the language model. Other configuration factors which we finetune include the layer of the language model embedding used (since studies have shown [47] that different layers of a language model encode different linguistic information within a sentence), choosing the token embedding (CLS vs mean), method of text preprocessing and which part of the text to select the 512 tokens from (e.g., first 512, last 512, first 256 and last 256).

C. Experimental Configuration

Since there is a variance in the model performance based on the weight initialization and data order, we report aggregated 10 fold cross-validation performance of the outer resampling loop, averaged over 10 seeds (Fig. 1). In our finetuning setup, we experimented with logistic regression, SVM and a multi-layer perceptron (MLP) with 50 hidden units and 'relu' non-linearity. The optimizer used was Adam [48] with a binary cross entropy loss function. We report the results of the best performing model in Table II. We also experimented with larger MLP architectures while finetuning, however, it resulted in no evident performance gain.

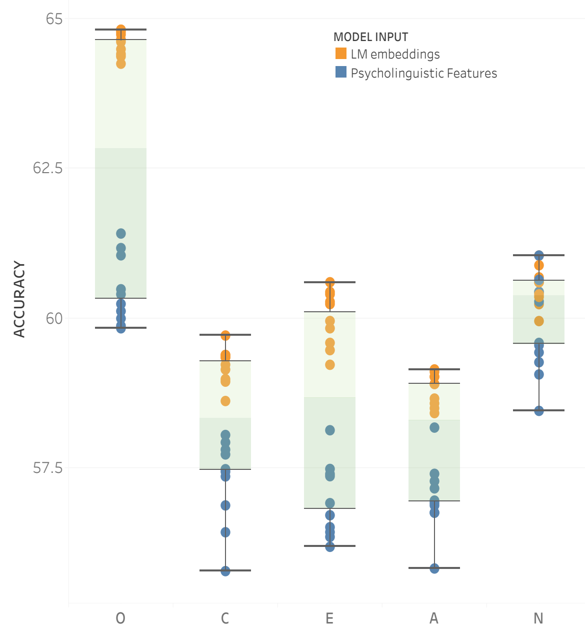

Fig. 1. Figure illustrating the variance of performance over different initialization seeds on the Essays dataset.

IV. Results

A. Predicting Personality with Language Data

We achieve state-of-the-art results on the Essays and Kaggle datasets. As can be seen in Table II, as expected, language model-based approaches far outperform the traditional closed vocab ones for personality prediction. However, we find that using a larger language model does not always result in higher performance. There is also a high variance in the predicted accuracy (Fig. 1) across runs with different model initialization seeds. The specific configuration of the language model used (as discussed in Section III-B2) also yields high variance in the model performance. For additional results, please refer to the open-science repository of the project.

B. The Importance of Psycholinguistic Features

Interpretable machine learning can be helpful to discover algorithmic biases and to discover invalid models (e.g., models using the wrong information for predictions [49]). Here, we

Fig. 2. Word cloud visualization of the most important psycholinguistic features driving personality trait prediction. Word size indicates the mean SHAP value.

analyzed the importance of the aforementioned psycholinguistic features using SHapley Additive exPlanations (SHAP) values [50], to quantify and to better understand the influence or predictors in a particular personality trait model.

In Fig. 2 and Tab. III we show results of the interpretable machine learning analysis. Openness was best predicted by the count of unique words, the number of 1st person singulars and the number of words referring to cognitive processes (e.g., cause, know, ought). Additionally, Fig. 2 suggests that the number of apostrophes was predictive for openness too. Openness is the personality trait dimension that is most closely related to intellect as well as diverse experiences, interests, and ideas [51]. Point-biserial correlations align with those reported in past research [20], [24], [25] and highlight linguistic characteristics with regard to high or low openness.

Predictions for the conscientiousness dimension were most influenced by the number of self-references in an essay, the number of causation words (e.g., because, effect), and the fraction of unique words in all words of an essay. Correlations between those features and conscientiousness are low, which could hint at those effects to be non-linear or interactive. Also, the correlation between conscientiousness and the number of causation words does not match with previous findings that reported negative [25] or no [24] linear association.

For the prediction of extraversion, the three most important linguistic features were the imageability rating (the degree to which words can evoke a clear mental image), the total number of sentences in an essay and the mean age of acquisition rating (estimate for when a word is on average learned as a kid) for the words in an essay.

Predictions for the dimension of agreeableness were most impacted by the total count of pronouns and swear words as well as by the mean number of syllables per word. The (negative) importance of swear words for the prediction of agreeableness has been reported before [20] and could point towards the tendency of agreeable people to act and to express themselves in a more polite and kinder manner [52].

Finally, for the prediction of the Big Five personality trait dimension neuroticism, our results suggest that the number of apostrophes, the LIWC anger value and the average number of

TABLE III

THIS TABLE SHOWS THE TOP THREE MOST INFLUENTIAL FEATURES IN THE ESSAYS TASK FOR BIG FIVER PERSONALITY TRAIT SCORE PREDICTION. ADDITIONALLY, THE POINT-BISERIAL CORRELATION LEVEL IS SHOWN. *P < .05. **P < .001, TWO-TAILED.

Trait	Top psycholinguistic feature for prediction	Feature Description	Point-biserial correlation coefficient
O	Dic	Count of unique words	−0.173
	I	Count of 1st person singulars	−0.136**
	Cogmech	Cognitive processes (LIWC)	0.033**
C	Self	Count of references to self	0.05*
	Cause	Causation (LIWC)	0.003
	type_token_ratio	Ratio of type of words (unique words) to the total number of words	−0.037
E	IMAG	Imageability rating (MRC)	−0.011
	sentences_per_paragraph	Number of sentences in the essay	−0.052*
	AOA	Age of acquisition : the age at which a word is typically learned	0.011
A	Pronoun	Count of pronouns	0.023
	Swear	Number of swear words	−0.117**
	Syllables	Average number of syllables per word	−0.016
N	Apostro	Count of apostrophe usage	0.045*
	Anger	Anger value from Mairesse	0.077**
	Syllables	Average number of syllables per word	0.035

syllables per word were most important. Additionally, Fig. 2 suggests words expressing anxiety and inhibition tendencies were important for the prediction of neuroticism.

In summary the results of our interpretable machine learning analysis partially align with past, associative findings from personality psychology and underline the expressiveness of language use for individual differences [53]. However, the findings also highlight that simple linear-associative analyses only poorly describe the relationship between linguistic features in text and personality traits.

Our integrative and interpretive approach to language-based personality prediction (bottom up an top down features) might help to close that gap between computational and theory driven approaches to personality science.

V. Discussion

In this paper, we proposed a novel deep learning-based model for language-based personality trait prediction. In this model we used traditional psycholinguistic features and language model embeddings as features. Additionally, we analyzed the contribution of individual psycholinguistic features on the final prediction of a personality trait. Our results show that language modeling features consistently beat conventional psycholinguistic features. Overall the BERT-base + MLP model dominated for the prediction of Big Five personality traits and BERT-large + MLP was mostly superior for the prediction of MBTI dimensions. The predictive performances of our models beat the current state-of-the-art on the Essays dataset by 1.6% and the Kaggle dataset by 1%. Furthermore, findings from our interpretable machine learning analysis partially align with past research in psychology [20], [24], [25].

A. Limitations & Outlook

While our results show improvements to other deep learning models using language data, there are a number of limitations that affect the present study. In psychometric personality trait assessments, personality is measured in continuous scores, yet the available benchmark datasets mostly provide personality traits scores in artificially binned form only. Future studies should aim to use datasets that provide continuous scores on personality traits. As common in language modeling we tested different model settings (e.g., token embeddings) on the complete dataset to identify optimal model settings for performance evaluation. This approach can lead to an overestimation of model performance. Hence, future studies should evaluate different model settings using a nested cross-validation approach [54].

Stachl et al. [49] talk further about the main challenges that researchers face when building, interpreting, and validating machine learning models for personality assessment. Another big drawback is that there are discrepancies between markers of self-assessed versus observed, and online versus offline personality. Besides, although our findings match prior evidence, the result might vary based on the analyzed socio-cultural group. Lewis [55] explored the diversity of individuals' behavior further. Finally, future works will investigate whether the application of SenticNet 6 [56] and the new Hourglass model [57] can improve the accuracy of personality prediction.

References

[1] D. J. Ozer and V. Benet-Martínez, "Personality and the Prediction of Consequential Outcomes," *Annual Review of Psychology*, vol. 57, no. 1, pp. 401–421, 2006.

[2] B. W. Roberts, N. R. Kuncel, R. Shiner, A. Caspi, and L. R. Goldberg, "The Power of Personality: The Comparative Validity of Personality Traits, Socioeconomic Status, and Cognitive Ability for Predicting Important Life Outcomes." *Perspectives on psychological science : a journal of the Association for Psychological Science*, vol. 2, no. 4, pp. 313–45, dec 2007.

[3] C. J. Soto, "How Replicable Are Links Between Personality Traits and Consequential Life Outcomes? The Life Outcomes of Personality Replication Project," *Psychological Science*, vol. 30, no. 5, pp. 711–727, may 2019. [Online]. Available: http://journals.sagepub.com/doi/10.1177/0956797619831612

[4] M. Ziegler, D. Bensch, U. Maaß, V. Schult, M. Vogel, and M. Bühner, "Big Five facets as predictor of job training performance: The role of specific job demands," *Learning and Individual Differences*, vol. 29, pp. 1–7, 2014.

[5] H. Yin, Y. Wang, Q. Li, W. Xu, Y. Yu, and T. Zhang, "A network-enhanced prediction method for automobile purchase classification using deep learning." in *PACIS*, 2018, p. 111.

[6] S. C. Matz, M. Kosinski, G. Nave, and D. J. Stillwell, "Psychological targeting as an effective approach to digital mass persuasion," *Proceedings of the national academy of sciences*, vol. 114, no. 48, pp. 12 714–12 719, 2017.

[7] H. Schneider, K. Schauer, C. Stachl, and A. Butz, "Your data, your vis: Personalizing personal data visualizations," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10515 LNCS. Springer, Cham, sep 2017, pp. 374–392. [Online]. Available: http://link.springer.com/10.1007/978-3-319-67687-6{_}25

[8] S. T. Völkel, R. Schödel, D. Buschek, C. Stachl, Q. Au, B. Bischl, M. B"uhner, and H. Hussmann, "Opportunities and Challenges of Utilizing Personality Traits for Personalization in HCI: Towards a shared perspective from HCI and Psychology," in *Personalized Human-Computer Interaction*. Oldenbourg, Germany: De Gruyter, 2019.

[9] C. C. Liem, M. Langer, A. Demetriou, A. M. Hiemstra, A. S. Wicaksana, M. P. Born, and C. J. König, "Psychology meets machine learning: Interdisciplinary perspectives on algorithmic job candidate screening," in *Explainable and Interpretable Models in Computer Vision and Machine Learning*. Springer, 2018, pp. 197–253.

[10] J. Maria Balmaceda, S. Schiaffino, and D. Godoy, "How do personality traits affect communication among users in online social networks?" *Online Information Review*, vol. 38, no. 1, pp. 136–153, 2014.

[11] E. Cambria, S. Poria, A. Gelbukh, and M. Thelwall, "Sentiment analysis is a big suitcase," *IEEE Intelligent Systems*, vol. 32, no. 6, pp. 74–80, 2017.

[12] R. J. Gregory, "The History of Psychological Testing," in *Psychological testing: History, principles, and applications.*, 7th ed. Pearson, 2013, ch. 2, pp. 32–58.

[13] O. P. John, L. P. Naumann, and C. J. Soto, "Paradigm shift to the integrative Big Five Trait taxonomy," *Handbook of personality: Theory and research*, pp. 114–158, 2008.

[14] R. R. McCrae, P. T. Costa, T. A. Martin, P. T. Costa Jr, and T. A. Martin, "The NEO–PI–3: A More Readable Revised NEO Personality Inventory," *Journal of Personality Assessment*, vol. 84, no. 3, pp. 261–270, 2005.

[15] I. Briggs Myers, "Introduction to type: A guide to understanding your results on the myers-briggs type indicator (revised by lk kirby & kd myers). palo alto," 1993.

[16] D. J. Pittenger, "Cautionary comments regarding the myers-briggs type indicator." *Consulting Psychology Journal: Practice and Research*, vol. 57, no. 3, p. 210, 2005.

[17] C. Luo, "The application of mbti theory in hiring sales staffs," in *The 19th International Conference on Industrial Engineering and Engineering Management*. Springer, 2013, pp. 703–709.

[18] C. Stachl, Q. Au, R. Schoedel, S. D. Gosling, G. M. Harari, D. Buschek, S. Theres, V. Olkel, T. Schuwerk, M. Oldemeier, T. Ullmann, H. Hussmann, B. Bischl, and M. Bühner, "Predicting personality from patterns of behavior collected with smartphones," *Proceedings of the National Academy of Sciences*, jul 2020. [Online]. Available: www.pnas.org/cgi/doi/10.1073/pnas.1920484117

[19] W. Youyou, M. Kosinski, and D. Stillwell, "Computer-based personality judgments are more accurate than those made by humans." *Proceedings of the National Academy of Sciences of the United States of America*, vol. 112, no. 4, pp. 1036–40, jan 2015.

[20] G. Park, H. A. Schwartz, J. C. Eichstaedt, M. L. Kern, M. Kosinski, D. J. Stillwell, L. H. Ungar, and M. E. Seligman, "Automatic personality assessment through social media language." *Journal of personality and social psychology*, vol. 108, no. 6, p. 934, 2015.

[21] S. C. Matz, R. E. Appel, and M. Kosinski, "Privacy in the age of psychological targeting," *Current Opinion in Psychology*, vol. 31, pp. 116–121, feb 2020.

[22] S. Freud, "On aphasia; a critical study." 1953.

[23] M. E. Ireland and J. W. Pennebaker, "Language style matching in writing: Synchrony in essays, correspondence, and poetry." *Journal of personality and social psychology*, vol. 99, no. 3, p. 549, 2010.

[24] H. A. Schwartz, J. C. Eichstaedt, M. L. Kern, L. Dziurzynski, S. M. Ramones, M. Agrawal, A. Shah, M. Kosinski, D. Stillwell, M. E. P. Seligman, and L. H. Ungar, "Personality, gender, and age in the language of social media: the open-vocabulary approach." *PloS one*, vol. 8, no. 9, p. e73791, jan 2013.

[25] T. Yarkoni, "Personality in 100,000 Words: A large-scale analysis of personality and word use among bloggers," *Journal of Research in Personality*, vol. 44, no. 3, pp. 363–373, jun 2010.

[26] J. W. Pennebaker and L. A. King, "Linguistic styles: Language use as an individual difference." *Journal of personality and social psychology*, vol. 77, no. 6, p. 1296, 1999.

[27] J. Pennebaker, M. Francis, and R. Booth, "Linguistic inquiry and word count (liwc): A computerized program," *Mahwah, NJ: Erlbraum*, 2001.

[28] A. Wang, A. Singh, J. Michael, F. Hill, O. Levy, and S. R. Bowman, "Glue: A multi-task benchmark and analysis platform for natural language understanding," *arXiv preprint arXiv:1804.07461*, 2018.

[29] P. Rajpurkar, J. Zhang, K. Lopyrev, and P. Liang, "Squad: 100,000+ questions for machine comprehension of text," *arXiv preprint arXiv:1606.05250*, 2016.

[30] E. C. Fieller, H. O. Hartley, and E. S. Pearson, "Tests for rank correlation coefficients. i," *Biometrika*, vol. 44, no. 3/4, pp. 470–481, 1957.

[31] F. Mairesse, M. A. Walker, M. R. Mehl, and R. K. Moore, "Using linguistic cues for the automatic recognition of personality in conversation and text," *Journal of artificial intelligence research*, vol. 30, pp. 457–500, 2007.

[32] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intelligent Systems and their applications*, vol. 13, no. 4, pp. 18–28, 1998.

[33] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781*, 2013.

[34] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532–1543.

[35] S. Poria, A. Gelbukh, B. Agarwal, E. Cambria, and N. Howard, "Common sense knowledge based personality recognition from text," in *Mexican International Conference on Artificial Intelligence*. Springer, 2013, pp. 484–496.

[36] N. Majumder, S. Poria, A. Gelbukh, and E. Cambria, "Deep learning-based document modeling for personality detection from text," *IEEE Intelligent Systems*, vol. 32, no. 2, pp. 74–79, 2017.

[37] A. Kazameini, S. Fatehi, Y. Mehta, S. Eetemadi, and E. Cambria, "Personality trait detection using bagged svm over bert word embedding ensembles," in *Proceedings of the ACL 2020 workshop on Widening NLP*. Association for Computational Linguistics, 2020.

[38] Y. Mehta, N. Majumder, A. Gelbukh, and E. Cambria, "Recent trends in deep learning based personality detection," *Artificial Intelligence Review*, vol. 53, pp. 2313–2339, 2020.

[39] M. Coltheart, "The mrc psycholinguistic database," *The Quarterly Journal of Experimental Psychology Section A*, vol. 33, no. 4, pp. 497–505, 1981.

[40] E. Cambria, S. Poria, D. Hazarika, and K. Kwok, "SenticNet 5: Discovering conceptual primitives for sentiment analysis by means of context embeddings," in *AAAI*, 2018, pp. 1795–1802.

[41] S. M. Mohammad and P. D. Turney, "Crowdsourcing a word-emotion association lexicon," vol. 29, no. 3, pp. 436–465, 2013.

[42] S. M. Mohammad, "Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 english words," in *Proceedings of The Annual Conference of the Association for Computational Linguistics (ACL)*, Melbourne, Australia, 2018.

[43] M. H. Amirhosseini and H. Kazemian, "Machine learning approach to personality type prediction based on the myers–briggs type indicator®," *Multimodal Technologies and Interaction*, vol. 4, no. 1, p. 9, 2020.

[44] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[45] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, "Albert: A lite bert for self-supervised learning of language representations," *arXiv preprint arXiv:1909.11942*, 2019.

[46] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.

[47] B. van Aken, B. Winter, A. Löser, and F. A. Gers, "How does bert answer questions? a layer-wise analysis of transformer representations," in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, 2019, pp. 1823–1832.

[48] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[49] C. Stachl, F. Pargent, S. Hilbert, G. M. Harari, R. Schoedel, S. Vaid, S. D. Gosling, and M. Bühner, "Personality Research and Assessment in the Era of Machine Learning," *European Journal of Personality*, p. per.2257, may 2020. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/per.2257

[50] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Curran Associates, Inc., 2017, pp. 4765–4774. [Online]. Available: http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf

[51] C. G. DeYoung, "Openness/intellect: A dimension of personality reflecting cognitive exploration." in *APA handbook of personality and social psychology, Volume 4: Personality processes and individual differences*, ser. APA handbooks in psychology., M. Mikulincer, P. R. Shaver, M. L. Cooper, and R. J. Larsen, Eds. Washington, DC, US: American Psychological Association, 2015, pp. 369–399.

[52] W. G. Graziano and R. M. Tobin, "Agreeableness." in *Handbook of individual differences in social behavior*, M. R. L. R. H. Hoyle, Ed. New York, NY, US: Guilford Press, 2009, pp. 46–61.

[53] R. L. Boyd and J. W. Pennebaker, "Language-based personality: a new approach to personality in a digital world," pp. 63–68, dec 2017.

[54] B. Bischl, O. Mersmann, H. Trautmann, and C. Weihs, "Resampling Methods for Meta-Model Validation with Recommendations for Evolutionary Computation," *Evolutionary Computation*, vol. 20, no. 2, pp. 249–275, 2012.

[55] R. Lewis, *When cultures collide*. Nicholas Brealey Publishing Boston, MA, 2010.

[56] E. Cambria, Y. Li, F. Xing, S. Poria, and K. Kwok, "SenticNet 6: Ensemble application of symbolic and subsymbolic AI for sentiment analysis," in *CIKM*, 2020.

[57] Y. Susanto, A. Livingstone, B. C. Ng, and E. Cambria, "The hourglass model revisited," *IEEE Intelligent Systems*, vol. 35, no. 5, 2020.