# Prompt-based Pre-trained Model for Personality and Interpersonal Reactivity Prediction

**Bin Li**[1][*] **Yixuan Weng**[2][*], **Qiya Song**[1][*], **Fuyan Ma**[1], **Bin Sun**[1], **Shutao Li**[1]

[1] College of Electrical and Information Engineering, Hunan University

[2] National Laboratory of Pattern Recognition Institute of Automation,

Chinese Academy Sciences, Beijing, 100190, China

{libincn, sqyunb, mafuyan, sunbin611, shutao_li}@hnu.edu.cn, wengsyx@gmail.com

## Abstract

This paper describes the LingJing team's method to the Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis (WASSA) 2022 shared task on Personality Prediction (PER) and Reactivity Index Prediction (IRI). In this paper, we adopt the prompt-based method with the pre-trained language model to accomplish these tasks. Specifically, the prompt is designed to provide knowledge of the extra personalized information for enhancing the pre-trained model. Data augmentation and model ensemble are adopted for obtaining better results. Extensive experiments are performed, which shows the effectiveness of the proposed method. On the final submission, our system achieves a Pearson Correlation Coefficient of **0.2301** and **0.2546** on Track 3 and Track 4 respectively. We ranked $1^{st}$ on both sub-tasks.

## 1 Introduction

Personality can be defined as a set of characteristics (e.g., age, income, and hobby), which can reflect the differences of individuals in thinking, emotions, and behaviours (Vora et al., 2020). The power of personality is worth exploring and pervades human lives everywhere (Beck and Jackson, 2022). Personality prediction is an interdisciplinary field spanning from psychology to computer science. However, people's personalities can't be directly observed and measured, but are expressed in activity patterns, and thus can be inferred in that way. Humans tend to covey their personalities through language because it is the most prominent way in such an Internet society. Meanwhile, written text is one of the most important appearances of language.

Consequently, the involvement of machine-learning-based methods in predicting the personality of individuals seems necessary. Over the past 20 years, much progress has been made in natural language processing (NLP), which is faced with a revolution. Especially, with the development of deep learning and transfer learning, automatically and accurately predicting the personality is an emerging topic in NLP. Classical text representation methods and pre-trained word representation approach both make the personality prediction research area more attractive and competitive. Even though how to compute and predict someone's personality based on texts is still an open question, attracting more and more researchers to focus on it.

The approaches of personality prediction have a long research history. Early in 2008, the Personae corpus (Luyckx and Daelemans, 2008) has already been proposed for predicting the personality from the text. The corpus is used for predicting the writer's personality traits that are reflected in writing style. Input representation is one of the most components in NLP. In recent years, the novel representation method has been the pre-trained word embeddings. Therefore, we mainly focus on the personality prediction approaches with the pre-trained word embeddings. Methods with the pre-trained embeddings are firstly based on Word2Vec (Mikolov et al., 2013b,a) and GloVe (Pennington et al., 2014). Surprisingly, Poria et al. (2013) propose to extract common sense knowledge and affective information to recognize personality from text, where they represent information in a directed graph. Despite the success of early word embedding models (e.g., Word2Vec), there are still practical problems. The first one is that previously unseen words make the model get into trouble. The second one is the overwhelmingly large parameters for a model to learn. Liu et al. (2016) propose a recurrent and compositional deep-learning-based model to address these issues.

Very recently, researchers start to explore large pre-trained models for NLP (Jawahar et al., 2019; He et al., 2020; Malkin et al., 2021). Kazameini et al. (2020) use the BERT language model (Devlin

---

[*]These authors contribute equally to this work.

et al., 2018) to extract contextualized word embeddings and achieve state-of-the-art performance for personality detection. Similarly, Transformer-MD (Yang et al., 2021) is proposed to put multiple posts together for representing the personality of each user. The context embeddings learned by large pre-trained models can effectively improve the performances and have theoretical advantages over traditional embedding methods. Therefore, we also adopt a pre-trained model named DeBERTa (He et al., 2020) to construct our personality prediction model. In this paper, we present our Prompt-based pre-trained network for personality prediction at Track 3 and Track 4 of WASSA-2022.

**Track 3:** Personality Prediction (PER), which consists in predicting the personality of the essay writer, knowing all his/her essays and the news article from which they reacted.

**Track 4:** Interpersonal Reactivity Index Prediction (IRI), which consists in predicting the personality of the essay writer, knowing all his/her essays and the news article from which they reacted.

## 2 Main Method

In this section, we will elaborate on our method in detail, including the model architecture, prompt design, regression optimization, data augmentation and model ensemble.

### 2.1 Model Architecture

The overall architecture of our method is shown in Figure 1, the DeBERTa pre-trained language model (He et al., 2020) is adopted as our backbone for personality and interpersonal reactivity prediction. We first input the text with a manually designed prompt to be tokenized with DeBERTa tokenizer. Then the input shall be encoded with the encoder through the self-attention (Vaswani et al., 2017) mechanism. Finally, the output is produced by the Linear layer for regression.

### 2.2 Prompt Design

Prompt Learning is considered to be the wise way for providing the pre-trained model with extra knowledge (Liu et al., 2021). For this reason, we manually design the prompt to extract relevant knowledge from the pre-trained model for personality prediction, which is presented as the fixed template, i.e., "A female, with fourth grade education, third race, 22 and income of 100000". Specifically, this persona information is mapped into the tokens
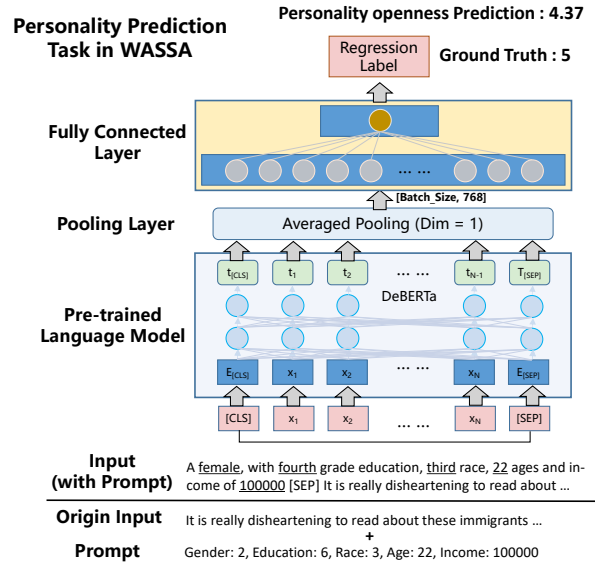


Figure 1: Overview of model architecture, where the sentence in origin input is concatenated with the prompt templates.

for providing more semantic information for the next regression task. We concatenate these fixed prompts with the origin input together for learning the joint representation in the pre-trained language model.

### 2.3 Regression Optimization

The personality and interpersonal reactivity prediction task is designed to regress the probable logits of different personality items. Given the training samples $D = \{(C^1, X^1), (C^2, X^2), \ldots, (C^D, X^D)\}$, where the $C^i$, $i = \{1, \ldots, n\}$, represents the author persona information with the corresponding stories collections $X^i$. The author persona information contains different other information items $C$, i.e., education and race, etc. We want to concatenate these texts together with the prompt learning, which aims to provide extra information for the personality prediction. The optimization function used the MSE function, which is shown as (1):

$$MSE(X, C, y) = \sum_{j=1}^{|\mathcal{D}|} \left( \text{Logits}\left(X^j, C^j\right) - y^j \right)^2 \tag{1}$$

where the Logits represents the logits output of prompt tuning from the pre-trained model, and $y^j$ is one author's personality item, $j \in [1, D]$.

We implement the above function with the opti-

mization of the following equation:

$$\mathcal{L} = -\sum_{k=1}^{|\mathcal{D}|} \log_\theta \left( y^k \mid X^k, C^k \right) \qquad (2)$$

where the $y^k$ represents each item personality label, and $\theta$ is the parameters of the pre-trained model.

## 2.4 Data Augmentation

Inspired by the work (Karimi et al., 2021), we consider the data augmentation with random punctuation marks, i.e., six punctuation marks in {".", ";", "?", ":", "!", ","}. The reason is that we want to ensure there is at least one inserted mark for more data from one author. Meanwhile, we do not want to insert too many punctuation marks as too much noise might hurt the model, especially for the personality prediction.

## 2.5 Model Ensemble

For different pre-trained models, the wise choice to improve the final results is to ensemble the pre-trained model (Zwanzig, 1960). As a result, we adopt the ensemble method to average the logits for the final prediction. Specifically, we implement the Bagging algorithm (Inoue and Kilian, 2008) for the personality and interpersonal reactivity prediction, which can effectively reduce the variance of the logits prediction by averaging the prediction bias produced from different models.

## 3 Experimental Setting

This section will subsequently present the emotion dataset, our experimental models, experimental settings, control of variables experiment.

## 3.1 Dataset

The shared task organizers supplied an extended dataset to participants used by (Buechel et al., 2018)[1]. The dataset includes essays between 300 and 800 characters with the Batson empathy, Personal Distress Scale, and other additional demographic and personality information. Among them, the person-level demographic information mainly contains age, gender, ethnicity, income and education level. The provided dataset of WASSA shared work contains 1860 training samples.

## 3.2 Implementation Details

In these tasks, we are mainly based on the hugging face framework[2] (Wolf et al., 2020). We add a randomly initialized linear layer after DeBERTa (He et al., 2021) to output the value of shape = [1]. We use the AdamW(Loshchilov and Hutter, 2018) optimizer and the learning rate is set to 8e-6 with the warm-up (He et al., 2016). The batch size is 12. We set the maximum length of 512, and delete the excess. Linear decay of learning rate and gradient clipping of 1e-4. Dropout (Srivastava et al., 2014) of 0.1 is applied to prevent overfitting. We implemented the code of training and reasoning based on PyTorch [3] (Paszke et al., 2019) in three NVIDIA A100 GPUs. All experiments select the best parameters in the valid set, and then report the score of the best model (valid set) in the test set.

We use the DeBERTa-v3-large[4] (He et al., 2021) as pre-trained model, and we Fine-tune the model. The DeBERTa-v3 model comes with 24 layers and a hidden size of 1036. This model uses a training framework similar to the ELECTRA (Clark et al., 2020) model, and sets up a generator and discriminator. The pre-training task of the discriminator is replaced with token detection (RTD). Finally, the RTD model is selected as the model. Compared with MLM, RTD training can bring more efficient training results.

## 4 Result and Discussion

This section describes our experiment results on Personality Prediction (PER) sub-task and the Interpersonal Reactivity Index Prediction (IRI) sub-task. Experiments were conducted with the development set as our test dataset, and the experimental results are shown in Table 1 and Table 2. The results of the final submissions are shown in Table 3.

## 4.1 Results for Track3

The results in Table 1 show the performance of our method with different components on the developments dataset. From Table 1, we can find that:

The method without ensemble component achieves 0.25788 AVG. Compared to the method without ensemble components, the combination of random punctuation and ensemble gains 0.0144

---

[1]Data and code are available at: https://github.com/wwbp/empathic_reactions

[2]https://github.com/huggingface/transformers
[3]https://pytorch.org
[4]microsoft/deberta-v3-large

| Methods | conscientiousness | extraversion | openess | talking | agreeableness | AVG |
|---|---|---|---|---|---|---|
| **Ours** | **0.2225** | **0.3194** | **0.2328** | **0.3815** | **0.4044** | **0.31212** |
| w/o random punctuation | 0.2197 | 0.3112 | 0.2246 | 0.3681 | 0.3640 | 0.29752 |
| w/o prompt | 0.1763 | 0.2841 | 0.2001 | 0.3148 | 0.3711 | 0.26928 |
| w/o ensemble | 0.1645 | 0.2467 | 0.1888 | 0.3201 | 0.3693 | 0.25788 |

Table 1: Results of the ablation study for Personality Prediction (PER). AVG represents the average of the Pearson correlations over Personality values. The best results are in bold.

| Methods | concern | distress | fantasy | stability | AVG |
|---|---|---|---|---|---|
| **Ours** | **0.3442** | **0.2871** | **0.2744** | **0.2108** | **0.2791** |
| w/o random punctuation | 0.3221 | 0.2881 | 0.2465 | 0.2049 | 0.2654 |
| w/o prompt | 0.3145 | 0.2645 | 0.2621 | 0.1893 | 0.2576 |
| w/o ensemble | 0.2953 | 0.2510 | 0.2346 | 0.1826 | 0.2408 |

Table 2: Results of the ablation study for Reactivity Index Prediction (IRI). AVG represents the average of the Pearson correlations over Personality values. The best results are in bold.

| Subtask | Pearson Correlations |
|---|---|
| PER (Track3) | 0.23006 |
| IRI (Track4) | 0.25460 |

Table 3: Results of our method on Track3 and Track4 respectively.

increase of AVG. The prompt and ensemble components are used in our method, which gains 0.02824 increase of AVG than W/o prompt. This shows that each component can improve the performance of our method. When the three components are used together, our model achieves state-of-the-art performance with 0.31212 AVG. On the final submissions shown in the Table 3, our method achieves a Pearson Correlation of 0.23006, which ranks **Top-1** in PER (Track3) sub-task.

### 4.2 Results for Track4

The results in Table 2 show the performance of each component for our approach on the development dataset of the Reactivity Index Prediction (IRI). From Table 2, we can find that:

Our method achieves 0.2791 AVG. Compared to the method without random punctuation, our method gains 0.0137 increase in AVG. The method without ensemble component only achieves 0.2408 AVG. These demonstrate the availability of introducing each component to our method. On the final submissions shown in the Table 3, our method achieves a Pearson Correlation of 0.25460, which ranks **TOP-1** in the IRI (Track4) sub-task.

## 5 Conclusion

In this paper, we describe our system to the Personality Prediction and Interpersonal Reactivity Index Prediction sub-tasks. We used the DeBERTa pre-trained language model as the backbone. The prompt is designed for providing the persona information to the pre-trained model. The data augmentation with random punctuation and model ensemble is adopted for better results. In the evaluation phase, our methods ranked **Top-1** on Track 3 and Track 4 respectively. In the future, we will focus on more effective prompt designing for performing the personality and interpersonal reactivity prediction.

## Acknowledgement

## References

Emorie D Beck and Joshua J Jackson. 2022. A mega-analysis of personality prediction: Robustness and boundary conditions. *Journal of Personality and Social Psychology*, 122(3):523.

Sven Buechel, Anneke Buffone, Barry Slaff, Lyle H. Ungar, and João Sedoc. 2018. Modeling empathy and distress in reaction to news stories. *arXiv: Computation and Language*.

Kevin Clark, Minh-Thang Luong, Quoc V. Le, and Christopher D. Manning. 2020. Electra: Pre-training text encoders as discriminators rather than generators. In *International Conference on Learning Representations*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778.

Pengcheng He, Xiaodong Liu, Jianfeng Gao, and Weizhu Chen. 2020. Deberta: Decoding-enhanced bert with disentangled attention. *arXiv preprint arXiv:2006.03654*.

Pengcheng He, Xiaodong Liu, Jianfeng Gao, and Weizhu Chen. 2021. Deberta: Decoding-enhanced bert with disentangled attention. In *International Conference on Learning Representations*.

Atsushi Inoue and Lutz Kilian. 2008. How useful is bagging in forecasting economic time series? a case study of us consumer price inflation. *Journal of the American Statistical Association*, 103(482):511–522.

Ganesh Jawahar, Benoît Sagot, and Djamé Seddah. 2019. What does bert learn about the structure of language? In *ACL 2019-57th Annual Meeting of the Association for Computational Linguistics*.

Akbar Karimi, Leonardo Rossi, and Andrea Prati. 2021. Aeda: An easier data augmentation technique for text classification. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 2748–2754.

Amirmohammad Kazameini, Samin Fatehi, Yash Mehta, Sauleh Eetemadi, and Erik Cambria. 2020. Personality trait detection using bagged svm over bert word embedding ensembles. *arXiv preprint arXiv:2010.01309*.

Fei Liu, Julien Perez, and Scott Nowson. 2016. A recurrent and compositional model for personality trait recognition from short texts. In *Proceedings of the Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media (PEOPLES)*, pages 20–29.

Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. 2021. Pretrain, prompt, and predict: A systematic survey of prompting methods in natural language processing. *arXiv preprint arXiv:2107.13586*.

Ilya Loshchilov and Frank Hutter. 2018. Decoupled weight decay regularization. In *International Conference on Learning Representations*.

Kim Luyckx and Walter Daelemans. 2008. Personae: a corpus for author and personality prediction from text. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*.

Nikolay Malkin, Sameera Lanka, Pranav Goel, Sudha Rao, and Nebojsa Jojic. 2021. Gpt perdetry test: Generating new meanings for new words. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5542–5553.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013a. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013b. Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, 26.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. 2019. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.

Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.

Soujanya Poria, Alexandar Gelbukh, Basant Agarwal, Erik Cambria, and Newton Howard. 2013. Common sense knowledge based personality recognition from text. In *Mexican International Conference on Artificial Intelligence*, pages 484–496. Springer.

Nitish Srivastava, Geoffrey E. Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15:1929–1958.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Hetal Vora, Mamta Bhamare, and Dr K Ashok Kumar. 2020. Personality prediction from social media text: An overview. *Int. J. Eng. Res*, 9(05):352–357.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. 2020. Transformers: State-of-the-art natural language processing. In

*Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.

Feifan Yang, Xiaojun Quan, Yunyi Yang, and Jianxing Yu. 2021. Multi-document transformer for personality detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 14221–14229.

Robert Zwanzig. 1960. Ensemble method in the theory of irreversibility. *The Journal of Chemical Physics*, 33(5):1338–1341.