CHAPTER FOUR

Text-based personality prediction using XLNet

Ashok Kumar Jayaraman[a], Gayathri Ananthakrishnan[b], Tina Esther Trueman[c], and Erik Cambria[d]

[a]Department of Information Science and Technology, Anna University, Chennai, India
[b]Department of Information Technology, VIT University, Vellore, India
[c]Department of Computer Science, University of the People, Pasadena, CA, United States
[d]School of Computer Science and Engineering, Nanyang Technological University, Singapore, Singapore

Contents

1. Introduction 50
2. Related works 51
3. Method 53
  3.1 Datasets 53
  3.2 XLNet 54
  3.3 XLNet fine-tuning 56
4. Results and discussion 57
5. Conclusion 62
Acknowledgments 62
References 62
About the authors 64

Abstract

Personality is a dynamic and organized set of characteristics that distinguish a person in thinking patterns, behaviors, emotions, and motivations based on biological and environmental factors. In particular, personality is a broad subject that is widely studied in various domains such as mental healthcare, web intelligence, and recommendation systems. Traditionally, researchers used psychology methods via psycholinguistic approaches (word counting in specific texts) to identify personality. Recently, social media data have been used for studies on personality. Psychologists and scientists are determining the personality of a person with the Myers–Briggs Type Indicator (MBTI) and Big Five model. Moreover, transformers-based models have shown better results in natural language processing tasks with context-dependent features. In this chapter, we propose a text-based personality prediction system using XLNet, which learns bidirectional context via factorization order and relative positional encoding. Experiments on two different gold-standard personality detection datasets show that the proposed model obtains up to 4% accuracy improvement.

Advances in Computers, Volume 132
ISSN 0065-2458
https://doi.org/10.1016/bs.adcom.2023.08.002

1. Introduction

Human personality is a colorful and complex system. It defines a person's feelings, behavior, and thinking patterns. In particular, personality is derived from an individual's experience and environmental factors. Therefore, the personality of an individual can change over time. However, adults relatively maintain their core personality traits during adulthood [1–3]. Earlier researchers used countless characteristics of an individual to determine personality. Nowadays, psychologists and scientists are determining the personality of a person with the Myers–Briggs Type Indicator (MBTI) and Big Five model. The MBTI is a personality type identification system that divides people into 16 distinct personality types across four axes, namely, Introversion (I)—Extroversion (E), Intuition (N)—Sensing (S), Thinking (T)—Feeling (F), and Judging (J)—Perceiving (P). First, I—E measures the outer vs inner world preference of an individual. Second, N—S measures the sensing vs impression patterns of an individual. Third, T—F determines objective principles and facts vs the emotional weights of an individual. Finally, J—P measures a planned and ordered life vs a spontaneous and flexible life of an individual.

The Big Five model is studied with five different personality types, namely, emotionality (or neuroticism), extroversion, agreeableness, conscientiousness, and openness (EXACO) [4, 5]. Neuroticism defines whether a person is sensitive, nervous, depressed, anxious, has negative feelings, and has self-doubt. Extroversion describes a person's talkativeness, outgoingness, and high energy. Agreeableness determines whether a person is cooperative, kind, trustworthy, polite, friendly, generous, and straightforward. Conscientiousness reflects whether a person is responsible, goal-directed, hard-working, and adheres strictly to norms and rules. Openness defines whether a person is curious about new ideas and new experiences. Some researchers use a six-factor model (HEXACO) for identifying personality. The six-factor model includes a new factor in addition to the Big Five model called honesty–humility. This personality trait reflects whether a person is moral, fair, sincere, and avoids greed [6].

Traditionally, researchers used psychology methods via psycholinguistic approaches (word counting in specific texts) to identify personality [7]. Psychology methods are broadly studied in two categories, namely, qualitative and quantitative. These categories are used in the form of case study, experiment, observational study, survey, and content analysis. In recent years, social media has become popular among internet users. They express their feelings and views in the form of audio, video, image, and text. These data become complex in nature to identify personality. Therefore, researchers used natural language processing (NLP), conventional machine learning, and deep learning methods to identify personality. They learn BoW (Bag of words) features and semantic context features. Moreover, the recurrent neural networks in deep learning capture a unidirectional context, i.e., from beginning to end and from end to beginning. However, Vaswani et al. [8] and Devlin et al. [9] introduced attention-based models such as transformer (encoder–decoder structure) and BERT (encoder structure) to capture bidirectional context with sequential parallelization. Later, Yang et al. [10] introduced a permutation language model (also called XLNet) to capture bidirectional context via factorization order and relative positional encoding. It has shown promising results over the BERT pretrained model. Thus, we propose a permutation language modeling for personality detection. This chapters contributes to the following.

- Addresses the MBTI and Big Five model for personality detection
- Employs a permutation language pretrained model
- Outperforms the personality detection task than the state-of-the-art models

The rest of this chapter is organized as follows: Section 2 describes the related works; Section 3 presents the text-based personality detection framework based on XLNet; Section 4 presents results and discussion; and finally, Section 5 offers concluding remarks.

2. Related works

Researchers studied the automated personality type detection system using various machine learning and deep learning models. The personality type detection system is used in various applications such as managerial style prediction [11], academic major selection [12], dentists' career choice prediction [13], college performance prediction system [14]. In particular, we discuss the text-based research works on MBTI and the Big Five model. Asghar et al. [15] investigated an attention-based BiLSTM for psychopath personality trait detection using social media text such as Facebook and Twitter. Their study indicated that the attention-based Bi-LSTM model achieves 85% accuracy for detecting psychopath and nonpsychopath. Stachl et al. [16] examined the Big Five personality model with six different behavioral information of an individual. This behavioral information is collected from smartphones via sensor and log data. Their study indicated that there are specific behavioral patterns in app usage, music consumption, mobility, day and night activity, and overall phone activity. Also, the authors suggested that there are benefits in terms of research and danger in terms of privacy implications. Amirhosseini and Kazemian [17] developed the MBTI-based automated personality type prediction system using the XGBoost algorithm. Specifically, the authors divided the 16 distinct classes into 4 binary classification tasks. They trained the binary classifier using TF-IDF features on each of the personality types separately. Their study indicated that the XGBoost model has shown reliability and better accuracy. Hernandez and Scott [18] performed the recurrent neural network to classify the MBTI personality type dataset. The authors divided the 16 MBTI classes into 4 binary tasks. Therefore, they trained four binary classifiers instead of a multiclass classifier. Their results indicated that the proposed RNN model achieves an average accuracy of 54.4% for the posts classification and 67.8% for user classification.

Ozer and Benet-Martinez [19] studied the personality characteristics and consequential outcomes using various meta-analyses. Their study claimed that the personality effects are influencing each of us while aggregating at the population level. Li et al. [20] proposed a multitask learning framework to predict emotional behavior and personality traits. Their study indicated that the convolutional neural network (CNN) achieves better performance on various personality and emotion datasets. Kazameini et al. [21] developed a bagged SVM model using contextualized word embedding for personality trait detection. The authors have broken the essay dataset into multiple chunks to extract maximum information. These subdocuments are associated with the same personality type. Later, they performed an SVM binary classifier to predict the corresponding personality trait. This method achieves 59.03% accuracy on average. Poria et al. [22] proposed a common-sense knowledge-based architecture to detect personality from the text. They achieved 63.6% average accuracy using the support vector machine (SVM). Majumder et al. [23] presented a CNN-based personality trait detection system. Their study indicated that CNN outperforms for all five personality traits with different configurations, and n-gram models showed no improvement. In summary, the existing researchers studied the personality type prediction system using a bag of word features, context-independent features, and context-dependent features. In this chapter, we propose the text-based personality detection system using permutation language modeling, where it supports context-dependent features via factorization order.

3. Method

In this section, we present the text-based personality prediction system using XLNet as shown in Fig. 1. We explain each part of this model as follows.

3.1 Datasets

We use two datasets for the personality identification system such as the MBTI dataset [24] and Essays dataset [25]. First, we use the MBTI dataset for the personality type identification system. This dataset groups each individual into 16 distinct personalities across four axes, namely, Introversion (I)—Extroversion (E), Intuition (N)—Sensing (S), Thinking (T)—Feeling (F),

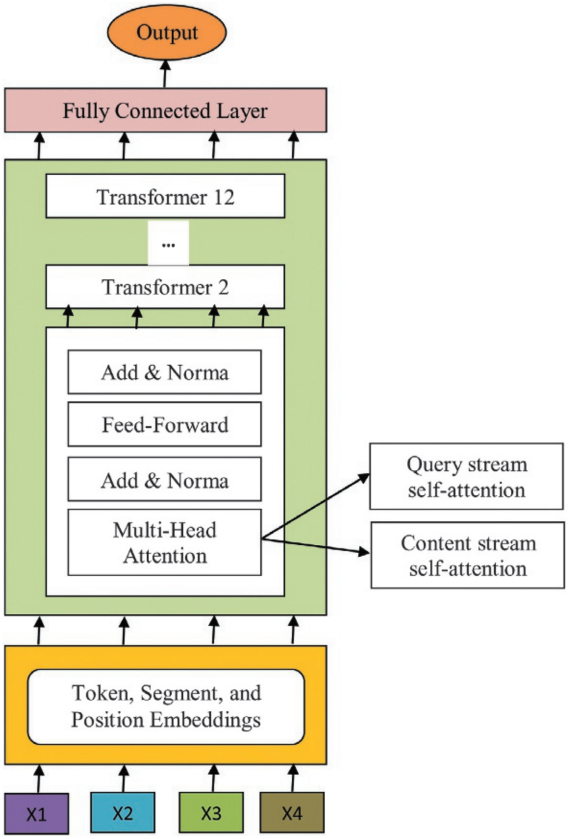

Fig. 1 The proposed model.

and Judging (J)—Perceiving (P). In particular, the MBTI dataset contains 8675 user posts associated with their personality type as in Table 1. Each personality type is encoded with four letters like INTJ (introversion, intuition, thinking, and judging). Second, we use the Essays dataset for the personality type identification system. This dataset contains 2467 anonymous essays associated with the Big Five personality measures such as emotionality (or neuroticism), extroversion, agreeableness, conscientiousness, and openness (EXACO) as in Table 2.

3.2 XLNet

XLNet is one of the latest pretrained models that achieve promising results in various NLP tasks such as document ranking, question answering, and sentiment analysis. Let $x = [x_1, x_2, \ldots, x_T]$ be the given input text with length T. Then, autoregressive language model ARLM maximizes the likelihood in a unidirectional way, i.e., from the forward or backward factorization order. This model fails to capture context information from both forward

Table 1 MBTI personality data distribution.

Personality	#instances	Personality	#instances
ENFJ	190	INFJ	1470
ENFP	675	INFP	1832
ENTJ	231	INTJ	1091
ENTP	685	INTP	1304
ESFJ	42	ISFJ	166
ESFP	48	ISFP	271
ESTJ	39	ISTJ	205
ESTP	89	ISTP	337

Table 2 Essays personality data distribution.

Personality	N Class	Y Class	#instances
Extroversion	1191	1276	2467
Neuroticism	1234	1233	2467
Agreeableness	1157	1310	2467
Conscientiousness	1214	1253	2467
Openness	1196	1271	2467

and backward directions. Therefore, ELMo [26] is introduced to capture context information from both forward and backward directions using a bidirectional language model. However, this model concatenates the contextual information from both directions, which are separately learned from the forward and backward directions. On the other hand, autoencoding language model learns contextual information from both forward and backward directions. Specifically, the BERT language model is designed to learn context from its surrounding information. This model predicts the masked words or tokens in the given input text. However, the BERT model has some critical aspects such as independence assumption between masked tokens, input noise, and context dependency. To address these critical aspects, Yang et al. [10] proposed a generalized autoregressive language model (XLNet). The XLNet uses factorization order and positional encoding to learn bidirectional context information. The objective function of this model is mathematically defined as in Eq. (1).

$$\max_{\theta} \; E_{z\sim Z_T}\left[\sum_{t=1}^{T} \log p_\theta(x_{z_t}|X_{z_{<t}})\right] \tag{1}$$

where Z_T denotes the set of all possible permutations of the input sequence with length T. z_t denotes the current $(t - th)$ element. $z_{<t}$ denotes the previous elements $(t - 1)$ of a permutation $z \in Z_T$. The objective function takes the previous elements as the input context for predicting the current element or token. In particular, the objective function considers the permutation-based factorization order rather than the input sequence order. Let $S = I$, *like*, *this*, *news* be the input text. Then, the permutation order of this input text is 4! Let $[3 \rightarrow 1 \rightarrow 2 \rightarrow 4]$ and $[1 \rightarrow 2 \rightarrow 4 \rightarrow 3]$ be the two sequence orders. Now, we predict the target element "this," which is computed as $P(this)$. This target element appears first in the first sequence order and last in the second sequence order. However, the first sequence order has no preceded elements to look over and the second sequence order looks at all preceded elements to compute the probability of target element as $P(this \mid I, like, news)$. Moreover, the transformer-based XLNet architecture uses two-stream self-attention mechanisms such as content stream and query stream for achieving this kind of permutation language modeling. First, the content stream self-attention mechanism encodes the context and content in the form of the standard hidden state representations as in Eq. (2). Second, the query stream self-attention mechanism encodes the context and positional information as in Eq. (3).

$$h_{z_t}^{(m)} \leftarrow Attention(Q = h_{z_t}^{(m-1)}, KV = h_{z_{\leq t}}^{(m-1)}; \theta) \tag{2}$$

$$g_{z_t}^{(m)} \leftarrow Attention(Q = g_{z_t}^{(m-1)}, KV = h_{z_{<t}}^{(m-1)}; \theta) \tag{3}$$

where Q, K, and V denotes the query, key, and value, and m denotes the number of attention layers.

In addition, XLNet reduces the optimization problem in the autoregressive language model using partial prediction. This predicts the last elements in a factorization order. Therefore, the log-likelihood of a target subsequence ($z_{>c}$) is maximized that conditioned on a nontarget subsequence ($z_{\leq c}$) as in Eq. (4). Here, c is the cutting point of a subsequence. ($Z_{>c}$) possesses the longest contextual information in the factorization order (z). Moreover, XLNet adopts the relative positional encoding and segment recurrence mechanism from Transformer-XL. These two techniques help XLNet to improve the long-term dependency of the given input text. Therefore, XLNet pretrained model is used as an effective model for NLP tasks.

$$\max_{\theta} \; E_{z \sim Z_T}\left[\log p_{\theta}(x_{z_{>c}} | X_{z_{\leq c}})\right] = E_{z \sim Z_T}\left[\sum_{t=c+1}^{|z|} \log p_{\theta}(x_{z_t} | X_{z_{<t}})\right] \tag{4}$$

In addition, XLNet incorporates two important techniques from Transformer-XL, namely, relative positional encoding and segment recurrence mechanism. These techniques help to improve the long-range dependency of the input sequence. Therefore, XLNet is used as an effective pretrained language model for NLP tasks.

3.3 XLNet fine-tuning

We use the XLNet pretrained model for identifying the personality of an individual from text. It is developed in two variants, namely, XLNet base model and XLNet large model. The XLNet base model represents 110M parameters with 12 transformer layers, 12 attention heads, and 768 hidden state units, and the XLNet large model represents 340M parameters with 24 transformer layers, 24 attention heads, and 1024 hidden state units. These architectures represent the same as BERT base and BERT large model. In particular, the BERT is build with the encoder structure of the transformer, and XLNet is build with the decoder structure of the transformer. The XLNet accepts [*CLS*, *A*, *SEP*, *B*, *SEP*] as the input format, where [*CLS*] and [*SEP*] are special elements or tokens to represent a classification token and sentence differentiation, and [*A*] and [*B*] are input segments. In the personality detection task, the given input sequence is aggregated by the token embedding, relative segment embedding, and relative position embedding. This aggregated input representation is fed to the transformer blocks. Then, an output layer with softmax activation function is added on the top of the XLNet transformers for calculating the output probabilities for each category.

4. Results and discussion

We experimented with the XLNet base model in Google Colab Pro on two different personality detection datasets, namely, the MBTI dataset and essays dataset. The MBTI dataset contains 8675 user posts, where each post is assigned with 1 of the 16 distinct personality types. We used the stratified random sampling method to divide the dataset into training (7026), validation (781), and testing (868) as in Table 3. Similarly, the essays dataset

Table 3 MBTI personality data split for training, validation, and testing.

Personality	Training	Validation	Testing
ENFJ	154	17	19
ENFP	546	61	68
ENTJ	187	21	23
ENTP	554	62	69
ESFJ	34	4	4
ESFP	39	4	5
ESTJ	31	4	4
ESTP	72	8	9
INFJ	1191	132	147
INFP	1484	165	183
INTJ	884	98	109
INTP	1057	117	130
ISFJ	134	15	17
ISFP	220	24	27
ISTJ	166	19	20
ISTP	273	30	34
Total	7026	781	868

Table 4 Essays personality data split for training, validation, and testing.

	Train (#1998)		Valid (#222)		Test (#247)		
Personality	**N**	**Y**	**N**	**Y**	**N**	**Y**	**Total**
Extroversion	965	1033	107	115	119	128	2467
Neuroticism	999	999	111	111	124	123	2467
Agreeableness	937	1061	104	118	116	131	2467
Conscientiousness	983	1015	109	113	122	125	2467
Openness	968	1030	108	114	120	127	2467

contains 2467 essays, where each essay is associated with the Big-Five personality type. This dataset is divided into training (1998), validation (222), and testing (247) for each personality type using the stratified random sampling as in Table 4. For instance, the extroversion personality type is divided into 80:10:10 for training (1998), validation (222), and testing (247). Specifically, we expanded the shorten texts to full text (e.g., "aren't" into "are not") using contraction map dictionary and removed punctuations except for the period, question mark, and exclamation mark. We then represent these datasets into the input tokens, segment tokens, and position tokens. These tokens are aggregated and fed into the proposed XLNet base pretrained model. In particular, we conduct a 16-class classification task on the MBTI dataset and five 2-class classification tasks on the essays dataset. For fine-tuning task, we used Adam optimizer with learning rate 2e-5, 1000 input sequence lengths, 8 epochs, 4 batch sizes, and 14,056 training steps for the MBTI dataset, and Adam optimizer with learning rate 2e-6, 1000 input sequence lengths, 8 epochs, 4 batch sizes, and 4000 training steps for the essays dataset.

Moreover, the XLNet base model is constructed with 110M parameters. It learns bidirectional context information via relative positional encoding and factorization order. We use the standard evaluation metrics such as precision (P), recall (R), and F1-score (F1) and their micro average, macro average, and weighted average [29] for computing the performance of the personality detection task. Table 5 shows the obtained validation and testing results for the MBTI dataset. This table indicates that the proposed fine-tuning model achieves 72% for micro precision, micro recall, and micro F1-score for personality detection tasks in the validation and testing datasets. Tables 6 and 7 show the five 2-class personality detection performances for validation and testing. In these tables, the openness personality type achieves 64% for micro precision, micro recall, and micro F1-score for both validation and testing datasets. It is higher than the other personality types.

Table 5 Performance of the MBTI dataset.

	Valid			Test		
Personality	**P**	**R**	**F1**	**P**	**R**	**F1**
ENFJ	0.82	0.82	0.82	0.65	0.79	0.71
ENFP	0.78	0.74	0.76	0.72	0.75	0.73
ENTJ	0.74	0.67	0.70	0.67	0.78	0.72
ENTP	0.64	0.68	0.66	0.76	0.70	0.73
ESFJ	0.67	1.00	0.80	1.00	0.75	0.86
ESFP	0.50	0.25	0.33	0.67	0.40	0.50
ESTJ	0.00	0.00	0.00	0.80	1.00	0.89
ESTP	1.00	0.50	0.67	0.80	0.44	0.57
INFJ	0.76	0.66	0.70	0.79	0.70	0.74
INFP	0.70	0.84	0.77	0.68	0.83	0.75
INTJ	0.69	0.73	0.71	0.76	0.65	0.70
INTP	0.68	0.70	0.69	0.71	0.75	0.73
ISFJ	0.71	0.67	0.69	0.74	0.82	0.78
ISFP	0.73	0.67	0.70	0.57	0.48	0.52
ISTJ	0.87	0.68	0.76	0.71	0.50	0.59
ISTP	0.89	0.57	0.69	0.76	0.65	0.70
Micro	0.72	0.72	0.72	0.72	0.72	0.72
Macro	0.70	0.64	0.65	0.74	0.69	0.70
Weighted	0.72	0.72	0.71	0.73	0.72	0.72

The result comparison for the MBTI dataset is shown in Table 8. In this table, Varma [27] performed various machine learning models in the ratio of 60:40 (training: testing) and 70:30. The authors indicated that the logistic regression model achieves 58.19% accuracy in the 60:40 ratio, and the logistic regression and XG Boost models achieve 57.12% accuracy in the 70:40 ratio. Hernandez et al. [18] performed LSTM network as five 2-class classifications on the MBTI dataset and achieved 54.40% accuracy. Uzsoy [30] performed the BERT pretrained model using TPU (Tensor Processing Unit). The authors achieved 68.10% accuracy with 1500 input sequence lengths. Our proposed XLNet base model achieves 72% accuracy using GPU. Then, the result comparison for the Big Five personality dataset

Table 6 Validation performance of the essays dataset.

Personality	**Extroversion**			**Neuroticism**			**Agreeableness**			**Conscientiousness**			**Openness**		
Class	**P**	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**
N	0.57	0.61	0.59	0.63	0.50	0.55	0.70	0.18	0.29	0.62	0.49	0.54	0.64	0.60	0.62
Y	0.61	0.57	0.59	0.58	0.70	0.64	0.56	0.93	0.70	0.59	0.71	0.64	0.64	0.68	0.66
Micro	0.59	0.59	0.59	0.60	0.60	0.60	0.58	0.58	0.58	0.60	0.60	0.60	0.64	0.64	0.64
Macro	0.59	0.59	0.59	0.60	0.60	0.59	0.63	0.56	0.50	0.60	0.60	0.59	0.64	0.64	0.64
Weighted	0.59	0.59	0.59	0.60	0.60	0.59	0.63	0.58	0.51	0.60	0.60	0.59	0.64	0.64	0.64

Table 7 Testing performance of the essays dataset.

Personality	**Extroversion**			**Neuroticism**			**Agreeableness**			**Conscientiousness**			**Openness**		
Class	**P**	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**	**P**	**R**	**F1**
N	0.51	0.66	0.61	0.60	0.54	0.57	0.69	0.21	0.32	0.63	0.52	0.57	0.62	0.64	0.63
Y	0.62	0.53	0.57	0.58	0.64	0.61	0.57	0.92	0.70	0.60	0.70	0.64	0.65	0.63	0.64
Micro	0.59	0.59	0.59	0.59	0.59	0.59	0.58	0.58	0.58	0.61	0.61	0.61	0.64	0.64	0.64
Macro	0.59	0.59	0.59	0.59	0.59	0.59	0.63	0.56	0.51	0.61	0.61	0.61	0.64	0.64	0.64
Weighted	0.60	0.59	0.59	0.59	0.59	0.59	0.62	0.58	0.52	0.61	0.61	0.61	0.64	0.64	0.64

Table 8 Result comparison for the MBTI dataset.

Authors	Model	Test
Varma (60:40) [27]	RandomForest	39.53
	XG Boost	57.87
	Gradient Descent	41.10
	Logistic Regression	58.19
	KNN	16.45
	SVM	35.52
Varma (70-30) [27]	RandomForest	37.87
	XG Boost	57.12
	Gradient Descent	45.13
	Logistic Regression	57.12
	KNN	16.96
	SVM	35.76
Hernandez et al. [18]	LSTM	54.40
Uzsoy [30]	BERT	68.10
Proposed	XLNet-Base-FT	72.24

Table 9 Result comparison for the essays dataset.

Personality	Majority baseline [21, 23]	Mairesse [23, 28]	CNN + Mairesse [21, 23]	BB-SVM [21]	Proposed	
					Valid	Test
Extroversion	51.72	55.13	58.09	59.30	59.01	59.10
Neuroticism	50.20	58.90	57.33	59.39	59.91	59.11
Agreeableness	53.10	55.35	56.71	56.52	58.11	58.30
Conscientiousness	50.79	55.28	56.71	57.84	59.91	61.13
Openness	51.52	59.57	61.13	62.09	64.41	63.56
Average	51.43	56.84	57.99	59.03	**60.27**	**60.24**

is shown in Table 9. Our proposed method outperforms than the majority baseline [21, 23], Mairesse [23, 28], CNN and Mairesse [21, 23], and BB-SVM [21] models. In the BB-SVM model, the authors had broken the essays into multiple subdocuments with 200 input sequence tokens. Therefore, we indicate that our proposed XLNet fine-tuning model achieves better results in both datasets.

5. Conclusion

In this chapter, we presented a text-based personality prediction system using permutation language modeling. This model captures bidirectional context-dependent features via factorization order and positional encoding. In particular, we expanded the shorten texts and removed punctuations except for the period, exclamation mark, and question mark. These help us to maintain the segment context. We then performed a 16-class personality detection system on the MBTI dataset and five 2-class personality detection systems on the essays dataset. We evaluated the proposed fine-tuning model using precision, recall, F1-score, and their macro average, micro average, and weighted average scores. Our results indicate that the proposed model achieves a 72% micro F1-score on the MBTI dataset and a 60% micro F1-score on the essays dataset. In future work, we desire to study the personality of an individual in cybercrime activity in a large dataset with gender and age group.

Acknowledgments

This work was supported by the University Grants Commission (UGC), Government of India under the National Doctoral Fellowship.

References

[1] Psychology Today website, Personality and its traits, https://www.psychologytoday.com/us/basics/personality, (accessed 05.06.21).
[2] M. Gjurković, J. Šnajder, Reddit: a gold mine for personality prediction, in: Proceedings of the Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media, 2018, pp. 87–97.
[3] A.H. Buss, Personality as traits, Am. Psychol. 44 (11) (1989) 1378.
[4] J.M. Digman, Personality structure: emergence of the five-factor model, Annu. Rev. Psychol. 41 (1) (1990) 417–440.
[5] Y. Mehta, N. Majumder, A. Gelbukh, E. Cambria, Recent trends in deep learning based personality detection, Artif. Intell. Rev. 53 (2020) 2313–2339.
[6] M.C. Ashton, K. Lee, M. Perugini, P. Szarota, R.E. De Vries, L. Di Blas, K. Boies, B. De Raad, A six-factor structure of personality-descriptive adjectives: solutions from psycholexical studies in seven languages, J. Pers. Soc. Psychol. 86 (2) (2004) 356.
[7] S. Štajner, S. Yenikent, A survey of automatic personality detection from texts, in: Proceedings of the 28th International Conference on Computational Linguistics, 2020, pp. 6284–6295.
[8] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, arXiv:1706.03762 (2017).
[9] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: pre-training of deep bidirectional transformers for language understanding, arXiv:1810.04805 (2018).
[10] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. Salakhutdinov, Q.V. Le, Xlnet: Generalized autoregressive pretraining for language understanding, arXiv:1906.08237 (2019).

[11] E.Y.D. Sari, K. Bashori, Predicting managerial styles: is the Myers-Briggs type indicator still useful? J. Educ. Learn. (EduLearn) 14 (4) (2020) 617–622.
[12] C.A. Pulver, K.R. Kelly, Incremental validity of the Myers-Briggs type indicator in predicting academic major selection of undecided university students, J. Career Assess. 16 (4) (2008) 441–455.
[13] T.G. Grandy, G.H. Westerman, R.A. Ocanto, C.G. Erskine, Predicting dentists'career choices using the Myers-Briggs type indicator, J. Am. Dent. Assoc. 127 (2) (1996) 253–258.
[14] L.J. Stricker, H. Schiffman, J. Ross, Prediction of college performance with the Myers-Briggs type indicator, Educ. Psychol. Meas. 25 (4) (1965) 1081–1095.
[15] J. Asghar, S. Akbar, M.Z. Asghar, B. Ahmad, M.S. Al-Rakhami, A. Gumaei, Detection and classification of psychopathic personality trait from social media text using deep learning model, Comput. Math. Methods Med. 2021 (2021) 1–10.
[16] C. Stachl, Q. Au, R. Schoedel, S.D. Gosling, G.M. Harari, D. Buschek, S.T. Völkel, T. Schuwerk, M. Oldemeier, T. Ullmann, Predicting personality from patterns of behavior collected with smartphones, Proc. Natl. Acad. Sci. U.S.A 117 (30) (2020) 17680–17687.
[17] M.H. Amirhosseini, H. Kazemian, Machine learning approach to personality type prediction based on the Myers-Briggs type indicator®, Multimodal Technol. Interact. 4 (1) (2020) 9.
[18] R.K. Hernandez, I. Scott, Predicting Myers-Briggs type indicator with text, in: 31st Conference on Neural Information Processing Systems (NIPS 2017), 2017.
[19] D.J. Ozer, V. Benet-Martinez, Personality and the prediction of consequential outcomes, Annu. Rev. Psychol. 57 (2006) 401–421.
[20] Y. Li, A. Kazameini, Y. Mehta, E. Cambria, Multitask Learning for Emotion and Personality Detection, arXiv:2101.02346 (2021).
[21] A. Kazameini, S. Fatehi, Y. Mehta, S. Eetemadi, E. Cambria, Personality trait detection using bagged SVM over BERT word embedding ensembles, in: Proceedings of WiNLP, 2020.
[22] S. Poria, A. Gelbukh, B. Agarwal, E. Cambria, N. Howard, Common sense knowledge based personality recognition from text, in: Mexican International Conference on Artificial Intelligence, Springer, 2013, pp. 484–496.
[23] N. Majumder, S. Poria, A. Gelbukh, E. Cambria, Deep learning-based document modeling for personality detection from text, IEEE Intell. Syst. 32 (2) (2017) 74–79.
[24] MBTI Kaggle dataset, https://www.kaggle.com/datasnaek/mbti-type, (accessed 27.05.21).
[25] J.W. Pennebaker, L.A. King, Linguistic styles: language use as an individual difference, J. Pers. Soc. Psychol. 77 (6) (1999) 1296.
[26] M.E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, L. Zettlemoyer, Deep contextualized word representations, arXiv:1802.05365 (2018).
[27] R. Varma, MBTI personality predictor using machine learning, 2021. https://www.kaggle.com/rajshreev/mbti-personality-predictor-using-machine-learning.
[28] F. Mairesse, M.A. Walker, M.R. Mehl, R.K. Moore, Using linguistic cues for the automatic recognition of personality in conversation and text, J. Artif. Intell. Res. 30 (2007) 457–500.
[29] J.A. Kumar, S. Abirami, A. Ghosh, T.E. Trueman, A C-LSTM with attention mechanism for question categorization, in: Symposium on Machine Learning and Metaheuristics Algorithms, and Applications, Springer, 2019, pp. 234–244.
[30] A.S. Uzsoy, Myers-Briggs types with tensorflow/BERT, 2020. https://www.kaggle.com/anasofiauzsoy/myers-briggs-types-with-tensorflow-bert.

About the authors

Ashok Kumar Jayaraman is a research scholar with the Department of Information Science and Technology, Anna University, Chennai. He received a Bachelor of Science in Mathematics from the University of Madras and a Master of Computer Applications from Anna University. Also, He has received UGC National Fellowship and Indian National Science Academy (INSA) Visiting Scientist Research Fellowship. His research interests include natural language processing, sentiment analysis, machine learning, deep learning, and social network analysis.

Gayathri Ananthakrishnan is currently working as an Assistant Professor (Senior) in the School of Computer Science Engineering and Information Systems (SCORE), VIT, Vellore. She has obtained her degree at Anna University. She has participated and presented at several National, International Conferences and Seminars. Her areas of interest include Wireless Network Security, Machine Learning, the Internet of Things (IoT), and Sentiment Analysis.

Tina Esther Trueman is an Assistant Professor with the Department of Computer Science, University of People, California. She obtained her M.E. in Computer Science and Engineering, Anna University, and obtained Ph.D. in Cloud data analytics from Anna University. She has published research papers in refereed journals and conferences. Her research interests include fuzzy logic systems, text mining, machine learning, deep learning, data mining, and data analytics.

Erik Cambria is a Professor at NTU, where he also holds the appointment of Provost Chair in Computer Science and Engineering. Prior to joining NTU, he worked at Microsoft Research Asia (Beijing) and HP Labs India (Bangalore) and earned his PhD through a joint program between the University of Stirling and MIT Media Lab. His research focuses on the ensemble application of symbolic and sub-symbolic AI to natural language processing tasks such as sentiment analysis, dialogue systems, and financial forecasting. Erik is recipient of many awards, e.g., the 2019 IEEE Outstanding Early Career Award, he was listed among the 2018 AI's 10 to Watch, and was featured in Forbes as one of the 5 People Building Our AI Future. He is Associate Editor of several journals, e.g., INFFUS, IEEE CIM, and KBS, Special Content Editor of FGCS, Department Editor of IEEE Intelligent Systems, and is involved in many international conferences as program chair and invited speaker.