

# Learning to Answer Psychological Questionnaire for Personality Detection

Feifan Yang, Tao Yang, Xiaojun Quan\*, Qinliang Su

School of Computer Science and Engineering, Sun Yat-sen University, China

{yangff6, yangt225}@mail2.sysu.edu.cn

{quanxj3, suqliang}@mail.sysu.edu.cn

## Abstract

Existing text-based personality detection research mostly relies on data-driven approaches to implicitly capture personality cues in online posts, lacking the guidance of psychological knowledge. Psychological questionnaire, which contains a series of dedicated questions highly related to personality traits, plays a critical role in self-report personality assessment. We argue that the posts created by a user contain critical contents that could help answer the questions in a questionnaire, resulting in an assessment of his personality by linking the texts and the questionnaire. To this end, we propose a new model named Psychological Questionnaire enhanced Network (PQ-Net) to guide personality detection by tracking critical information in texts with a questionnaire. Specifically, PQ-Net contains two streams: a context stream to encode each piece of text into a contextual text representation, and a questionnaire stream to capture relevant information in the contextual text representation to generate potential answer representations for a questionnaire. The potential answer representations are used to enhance the contextual text representation and to benefit personality prediction. Experimental results on two datasets demonstrate the superiority of PQ-Net in capturing useful cues from the posts for personality detection.

## 1 Introduction

As a psychological conception, personality aims to explain human behaviors in terms of a few stable and measurable individual characteristics (Vinciarrelli and Mohammadi, 2014). The study of personality is fundamental to psychology, and personality detection (Xue et al., 2018) has benefited many applications such as dialogue systems (Zheng et al., 2019), recommendation systems (Yang and Huang, 2019), and suicide risk assessment (Matero et al., 2019). Canonical approaches to personality test are

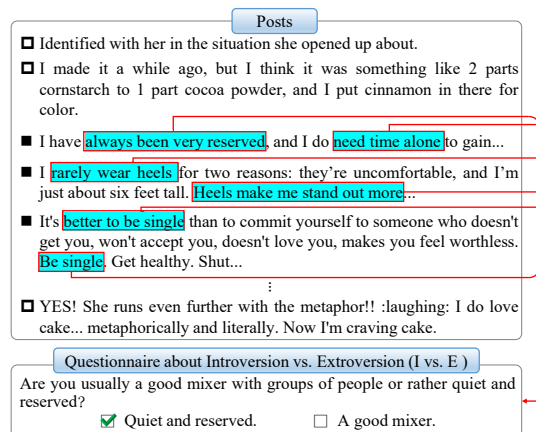


Figure 1: An example to show that certain online contents (highlighted) created by user can be used to answer the questions of a questionnaire. Highlighted contents strongly indicate that the right choice is 'Quiet and reserved'.

generally based on questionnaires elaborately designed by psychologists, yet the cost and scalability issues make them less practical in cyberspace (Nie et al., 2014; Aung and Myint, 2019).

Recent years have witnessed an increasing interest in automatically identifying one's personality traits based on her/his social media posts (Sorokowska et al., 2016; Imran et al., 2018; Tadesse et al., 2018b; Dandannavar et al., 2020). To encode the input posts and obtain their context representations, most of these methods employ deep learning models such as LSTMs (Tandera et al., 2017), CNNs (Xue et al., 2018) and pre-trained language models (PTMs) (Jiang et al., 2020; Yang et al., 2021a,b). They generally rely on the models to capture potential personality cues implicitly from the texts in a data-driven manner, without any guidance of psychological domain knowledge. As a result, the performance of these models is largely limited by the availability of training data and the learning capability of models.

We observe from real data that the posts created

\*Corresponding author.

by a user contain some critical contents that could help answer the questions in a questionnaire. As the example shows in Figure 1, there are a set of posts from a user and a question “Are you usually a good mixer with groups of people or rather quiet and reserved?” from an MBTI (Briggs and Myers, 1977) questionnaire. The question is also associated with two choices “Quite and reserved.” and “A good mixer.”, which are intended to investigate whether the user’s personality trait is introversive (the former) or extroversive (the latter). From the posts, we can see that the contents “always been very reserved”, “need time alone” and “better to be single” strongly indicate that the user’s personality trait is introversive. Therefore, we argue that it is possible to utilize the questionnaire, which contains questions that are highly related to personality traits, to guide a model to capture critical information in the posts for personality detection.

For this purpose, we propose a new model named Psychological Questionnaire enhanced Network (PQ-Net) for text-based personality detection. Specifically, PQ-Net consists of two streams: a context stream and a questionnaire stream. For the context stream, a PTM-based encoder is employed to encode each post and create its contextual representation. For the questionnaire stream, it first encodes each question by a question encoder and each candidate answer by a choice encoder, and then employs a cross-attention mechanism with supervision to enable the model to learn a potential answer representation for each question by choosing the correct answer based on the post representations. We then concatenate the post representations and the potential answer representations to predict the user’s personality traits. Under the guidance of the questionnaire, our PQ-Net is able to capture personality-related cues from the posts in an explicit manner rather than learning them implicitly.

Extensive experiments on the Kaggle and Pandora datasets show that PQ-Net consistently outperforms existing competitors with superior performance. Further analyses also demonstrate that the questionnaire and the two-streams structure all play a crucial role in PQ-Net, and that the user representations enhanced by PQ-Net are more inductive and distinguishable in comparison to the baselines. Lastly, we show that the cues obtained by PQ-Net are more interpretable for personality detection.

The contributions of this paper are threefold:

- This is the first work to introduce a traditional

psychological questionnaire into automatic personality detection, offering a new perspective of utilizing psychological knowledge.

- We propose a novel model to track critical information in posts with a questionnaire and provides an explicit way of identifying relevant cues in posts for personality detection.
- We demonstrate on two datasets that PQ-Net can effectively capture personality-relevant cues in posts and yield superior performance.

## 2 Methodology

### 2.1 Task Definition

The personality detection task studied in this paper can be formally defined as follows. Given a set  $P = \{p_i\}_{i=1}^n$  of social media posts from a user, where  $p_i = [w_{i,1}^p, w_{i,2}^p, \dots, w_{i,l_p}^p]$  represents the  $i$ -th post with  $l_p$  words. Consider an extra personality-related psychological questionnaire  $Q = \{(q_j, \{c_{j,k}\}_{k=1}^r)\}_{j=1}^m$  with  $m$  questions, where each question  $q_j = [w_{j,1}^q, w_{j,2}^q, \dots, w_{j,l_q}^q]$  is associated with  $r$  choices  $\{c_{j,k}\}_{k=1}^r$ . We use  $c_{j,k} = [w_{j,k,1}^c, w_{j,k,2}^c, \dots, w_{j,k,l_c}^c]$  to represent the  $k$ -th choice for question  $q_j$ . The objective of this task is to predict the personality traits  $Y = \{y^{(1)}, y^{(2)}, \dots, y^{(T)}\}$  of the user along  $T$  dimensions based on posts  $P$  and questionnaire  $Q$ .

### 2.2 Architecture

The overall architecture of our PQ-Net is demonstrated in Figure 2, which mainly comprises two streams: a context stream and a questionnaire stream. The context stream aims to encode each post by a post encoder to obtain its contextual representation (i.e., implicit cues). The questionnaire stream first encodes each question by a question encoder and each choice by a choice encoder. Then, it performs cross attention to capture key information in the contextual representations that can help “answer” the questions of the questionnaire, resulting in a potential answer representation for each question (i.e., explicit cues). The potential representations for all the questions are split into different categories according to their correspondences with the personality traits. Finally, the averaged contextual representation and the averaged answer representation in each category are concatenated as the enhanced representation to predict each personality

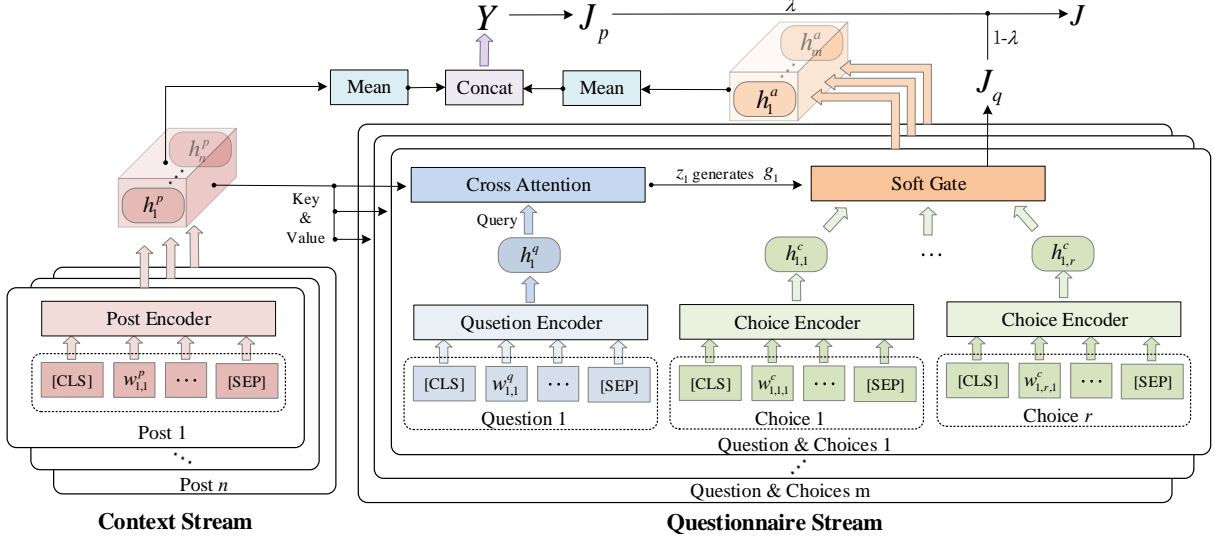


Figure 2: Overall architecture of our PQ-Net which comprises a context stream (left) to encode each post and create its contextual representation, and a questionnaire stream (right) to track question-relevant information in contextual representations and generate potential answer representations used to enhance contextual representations.

trait. In the following subsections, we introduce these two streams in detail, respectively.

### 2.2.1 Context Stream

As shown in the left part of Figure 2, the input of the context stream is a set of posts  $P = \{p_i\}_{i=1}^n$ . To obtain the representation of each post, BERT (Devlin et al., 2019) is employed as the encoder, while other PLMs apply equally. Formally, the contextual representation  $h_i^p$  of the  $i$ -th post is obtained by:

$$h_i^p = \text{BERT}^p \left( \text{CLS}, w_{i,1}^p, \dots, w_{i,l_p}^p, \text{SEP} \right). \quad (1)$$

where CLS and SEP are special tokens which represent the start and end of an input sequence, respectively.  $\text{BERT}^p(\cdot)$  denotes the final hidden state of the CLS token of BERT, which is commonly used as the abstract representation of a sequence. We apply the post encoder to encode each post and correspondingly obtain a set of contextual representations  $\mathbf{h}^p = [h_1^p, h_2^p, \dots, h_n^p] \in \mathbb{R}^{n \times d}$ , where  $d$  is the dimension of each representation.

### 2.2.2 Questionnaire Stream

As shown in the right part of Figure 2, we first encode each question via a question encoder and its choices via a choice encoder to obtain their contextual representations. In this study, the question encoder and the choice encoder are allowed to share the same pre-trained BERT parameters by considering their relatedness. Formally, similar to Eq. (1), we obtain the abstract representations of the  $j$ -th

question  $h_j^q$  and its corresponding  $k$ -th choice  $h_{j,k}^c$  as in Eq. (2) and Eq. (3), respectively.

$$h_j^q = \text{BERT}^q \left( \text{CLS}, w_{j,1}^q, \dots, w_{j,l_q}^q, \text{SEP} \right), \quad (2)$$

$$h_{j,k}^c = \text{BERT}^q \left( \text{CLS}, w_{j,k,1}^c, \dots, w_{j,k,l_c}^c, \text{SEP} \right), \quad (3)$$

We then apply a cross-attention mechanism (Vaswani et al., 2017) to capture critical information in the post representations by trying to “answer” the questions in the questionnaire. Specifically, the  $j$ -th question representation  $h_j^q$  is used as the query and the post representations  $\mathbf{h}^p$  are used as the key and value. Then, the question-aware post representation  $z_j$  for the  $j$ -th question is obtained by:

$$z_j = \parallel_{s=1}^S \sigma \left( \frac{h_j^q \mathbf{W}_s^Q (\mathbf{h}^p \mathbf{W}_s^K)^T}{\sqrt{d_k}} \right) (\mathbf{h}^p \mathbf{W}_s^V), \quad (4)$$

where  $S$  is the number of attention heads,  $d_k = \frac{d}{S}$  is the hidden size of each head, and  $\sigma$  is the softmax function.  $\mathbf{W}_s^Q \in \mathbb{R}^{d \times d_k}$ ,  $\mathbf{W}_s^K \in \mathbb{R}^{d \times d_k}$  and  $\mathbf{W}_s^V \in \mathbb{R}^{d \times d_k}$  are the  $s$ -th head linear transformations for query, key and value, respectively.  $\frac{1}{\sqrt{d_k}}$  is the scaling factor of attention weights, and  $\parallel$  is the concatenation operation along each head. Eq. (4) demonstrates how we track question-relevant information from the posts one by one and aggregate them through the attention weights. Then,  $z_j \in \mathbb{R}^{1 \times d}$  is used to predict the possibility of each choice being the answer of the  $j$ -th question:

$$g_j = \text{Softmax} (z_j \mathbf{W}^G + b^G), \quad (5)$$

where  $\mathbf{W}^G \in \mathbb{R}^{d \times r}$  and  $b^G \in \mathbb{R}^{1 \times r}$  are the learnable parameters of an affine transformation that converts  $z_j$  into  $r$  dimensions.

Note that each question in the questionnaire typically focuses on a certain personality trait, and thus its choices directly reflect the tendency of this personality trait. In other words, the preferred choice of each question can be inferred from the user’s personality traits during training, which provides additional supervision signals for the model to predict the choices correctly:

$$J_q = \frac{1}{m} \sum_{j=1}^m -\hat{g}_j \log g_j, \quad (6)$$

where  $\hat{g}_j$  is the preferred answer of the  $j$ -th question. Once we obtain the possibility of each choice, we then calculate the potential answer representation of the  $j$ -th question in a soft approach as follows:

$$h_j^a = \sum_{k=1}^r g_{j,k} h_{j,k}^c. \quad (7)$$

As a result, we obtain all the potential answer representations  $\mathbf{h}^a = [h_1^a, h_2^a, \dots, h_m^a] \in \mathbb{R}^{m \times d}$ .

### 2.3 Classification & Objective

Since each question in the questionnaire focuses on a specific personality trait, we divide the potential answer representations  $\mathbf{h}^a$  into  $T$  groups according to their correspondences with the personality traits. Formally, the  $t$ -th trait-specific answer representations are represented as follow:

$$\mathbf{h}^{a(t)} = \{h_j^a | q_j \in \text{trait}^{(t)}\}_{j=1}^m, \quad (8)$$

For each personality trait, the averaged post representation and the average trait-specific answer representation are concatenated to produce the final representation  $u^{(t)}$ :

$$u^{(t)} = (\text{mean}(\mathbf{h}^p) \parallel \text{mean}(\mathbf{h}^{a(t)})) \in \mathbb{R}^{1 \times 2d}, \quad (9)$$

Then,  $T$  softmax-normalized linear transformations are used to predict the probability on each personality trait, respectively. Concretely, for the  $t$ -th trait, we calculate:

$$p(y^{(t)}) = \text{Softmax}(u^{(t)} \mathbf{W}_u^{(t)} + b_u^{(t)}), \quad (10)$$

where  $\mathbf{W}_u^{(t)}$  is a trainable weight matrix and  $b_u^{(t)}$  is the bias term. The objective function of personality detection is defined as follows:

$$J_p = \frac{1}{T} \sum_{t=1}^T -y^{(t)} \log p(y^{(t)}), \quad (11)$$

where  $y^{(t)}$  denotes the true label of the  $t$ -th trait.

Finally, the tasks of questionnaire answering and personality detection are jointly trained with their functions linearly combined as follow:

$$J = \lambda \cdot J_p + (1 - \lambda) \cdot J_q. \quad (12)$$

where  $\lambda \in (0, 1)$  is a tunable coefficient.

## 3 Experiments

In this section, we first introduce the details of the personality benchmarks, questionnaire and baseline models adopted in our study, and then report and discuss our experimental results.

### 3.1 Datasets

Big Five and MBTI are two widely used personality frameworks in the fields of computational linguistics and natural language processing (Stajner and Yenikent, 2020). Presently, Big Five tends to be more reasonable to measure personality in psychology. However, most of the works (Xue et al., 2018; Lynn et al., 2020) on personality detection with Big Five are conducted on the unreleased myPersonality dataset, lacking a publicly accessible dataset based on social media with Big Five in the community (Gjurković et al., 2020). The studies (McCrae and Costa Jr, 1989; Costa Jr and McCrae, 1992) have shown that the Big Five and MBTI factions are correlated.

Therefore, we employ two available MBTI benchmarks of sufficient size, Kaggle<sup>1</sup> and Pandora<sup>2</sup>, to evaluate our method. While the former was collected from PersonalityCafe<sup>3</sup>, with 40-50 posts available for each user, the latter was collected from Reddit<sup>4</sup> with dozens of to hundreds

Dataset	Traits	Train(60%)	Validation(20%)	Test(20%)
Kaggle	<i>I</i> vs. <i>E</i>	4011 vs. 1194	1326 vs. 409	1339 vs. 396
	<i>S</i> vs. <i>N</i>	727 vs. 4478	222 vs. 1513	248 vs. 1487
	<i>T</i> vs. <i>F</i>	2410 vs. 2795	791 vs. 944	780 vs. 955
	<i>P</i> vs. <i>J</i>	3096 vs. 2109	1063 vs. 672	1082 vs. 653
Pandora	<i>I</i> vs. <i>E</i>	4278 vs. 1162	1427 vs. 386	1437 vs. 377
	<i>S</i> vs. <i>N</i>	610 vs. 4830	208 vs. 1605	210 vs. 1604
	<i>T</i> vs. <i>F</i>	3549 vs. 1891	1120 vs. 693	1182 vs. 632
	<i>P</i> vs. <i>J</i>	3211 vs. 2229	1043 vs. 770	1056 vs. 758

Table 1: Statistics of the number of samples under different categories in the two datasets.

<sup>1</sup><https://www.kaggle.com/datasnaek/mbti-type>

<sup>2</sup><https://psy.takelab.fer.hr/datasets>

<sup>3</sup><https://www.personalitycafe.com>

<sup>4</sup><https://www.reddit.com>



Methods	Kaggle					Pandora				
	<i>I/E</i>	<i>S/N</i>	<i>T/F</i>	<i>P/J</i>	Average	<i>I/E</i>	<i>S/N</i>	<i>T/F</i>	<i>P/J</i>	Average
SVM	53.34	47.75	76.72	63.03	60.21	44.74	46.92	64.62	56.32	53.15
XGBoost	56.67	52.85	75.42	65.94	62.72	45.99	48.93	63.51	55.55	53.50
LSTM <sub>Glove</sub>	57.82	57.87	69.97	57.01	60.67	48.01	52.01	63.48	56.21	54.93
BERT <sub>fine-tune</sub>	64.65	57.12	77.95	65.25	66.24	56.60	48.71	64.70	56.07	56.52
AttRCNN	59.74	64.08	78.77	66.44	67.26	48.55	<b>56.19</b>	64.39	57.26	56.60
SN-Attn	65.43	62.15	78.05	63.92	67.39	56.98	54.78	60.95	54.81	56.88
PQ-Net	<b>68.94</b>	<b>67.65</b>	<b>79.12</b>	<b>69.57</b>	<b>71.32</b>	<b>57.07</b>	55.26	<b>65.64</b>	<b>58.74</b>	<b>59.18</b>

Table 2: Overall results of different models in Macro-F1(%), where the best results are shown in bold.

of posts included for each user. The labels of the datasets are annotated based on self-diagnosis MBTI results (Gjurković and Šnajder, 2018), the taxonomy of which divides personality into four traits, each containing two possible values (Keh et al., 2019): *Introversion vs Extroversion (I vs E)*, *Sensing vs iNtuition (S vs N)*, *Think vs Feeling (T vs F)*, and *Perception vs Judging (P vs J)*. The statistics of the two datasets are shown in Table 1. Since the distribution of labels is imbalanced, we use the Macro-F1 metric for a more accurate evaluation.

### 3.2 Questionnaire

As shown in Appendix A, 26 personality-related questions are defined in the questionnaire<sup>5</sup>, each focusing on one of four MBTI personality traits with two choices reflecting the tendency of this personality trait. For example, the question “*Are you usually a good mixer with groups of people or rather quiet and reserved?*” focuses on the *I/E* trait, and the choices “*Quiet and reserved.*” and “*A good mixer.*” correspond to the *I* and *E* categories in this trait, respectively. Based on the ground truth personality labels, we can easily infer the preferred answer to each question for a user, which is treated as an extra supervision signal for training PQ-Net in Eq. (6). The exact number of questions for the *I/E*, *S/N*, *T/F* and *P/J* traits are 8, 7, 3 and 8, respectively.

### 3.3 Baselines

To make a comprehensive evaluation of our model, we employ the following models as baselines:

**SVM** (Cui and Qi, 2017) and **XGBoost** (Tadesse et al., 2018a) are respectively utilized as the classifier based on TF-IDF features.

**LSTM<sub>Glove</sub>** (Tandera et al., 2017) encodes each

post with Bi-LSTM and Glove (Pennington et al., 2014) word embedding and uses the averaged post representation as the user representation.

**BERT<sub>fine-tune</sub>** (Keh et al., 2019) is similar to LSTM<sub>Glove</sub> but encodes each post with the pre-trained BERT (Devlin et al., 2019).

**AttRCNN** (Xue et al., 2018) is the latest model that exploits LIWC psycholinguistic features.<sup>6</sup>

**SN-Attn** (Lynn et al., 2020) employs a hierarchical network with both word- and post-level attention.<sup>6</sup>

### 3.4 Implementation Details

We use Pytorch (Paszke et al., 2019) to implement our PQ-Net on four 2080Ti GPU cards. As the previous study (Hernandez and Knight, 2017), we set the maximum number of posts per user to 50 and the maximum length of each post to 70. For the questionnaire, we set the maximum length to 43/21 for each question/choice. For BERT, we use the bert-base-uncased (Devlin et al., 2019) to initialize. For training, we use the Adam (Kingma and Ba, 2014) optimizer with a mini-batch size of 32, a dropout rate of 0.2, and a learning rate of 2e-5/1e-3 for pre-trained/non-pretrained modules. For the coefficient  $\lambda$  in Eq. (12), we search it in the range of 0.1 to 0.9 with a step of 0.1 and set it to 0.7 eventually. During training, we use the early-stopping strategy for 5 consecutive epochs on validation set and report the final performance on test set.

### 3.5 Overall Results

The overall results are shown in Table 2, from which three observations can be noted. First, our PQ-Net consistently outperforms the other models on the two benchmarks. Particularly, on Kaggle, PQ-Net outperforms the latest state-of-the-art model (SN-Attn) and the basic pre-trained encoder (BERT<sub>fine-tune</sub>) they relies on by 3.93 and 5.08 in

<sup>5</sup><https://wedgworthleadership.com/wp-content/uploads/2016/08/Myers-Briggs-Personality-Test.pdf>

<sup>6</sup>We use BERT as the post encoder for a fair comparison.

Methods	Macro-F1 ( $\Delta\%$ )	
	Kaggle	Pandora
PQ-Net	<b>71.32</b>	<b>59.18</b>
w/o question	69.18 ( $\downarrow$ 2.14)	58.69 ( $\downarrow$ 0.49)
w/o choice	68.63 ( $\downarrow$ 2.69)	57.51 ( $\downarrow$ 1.67)
w/o soft gate	68.98 ( $\downarrow$ 2.34)	58.02 ( $\downarrow$ 1.16)
w/o $J_q$	68.74 ( $\downarrow$ 2.58)	57.28 ( $\downarrow$ 1.90)
w/o cross attention	69.73 ( $\downarrow$ 1.59)	58.46 ( $\downarrow$ 0.72)
w/o questionnaire stream	68.41 ( $\downarrow$ 2.91)	57.45 ( $\downarrow$ 1.73)
w/o context stream	70.04 ( $\downarrow$ 1.28)	58.23 ( $\downarrow$ 0.95)
w/o special encoder	70.65 ( $\downarrow$ 0.67)	58.16 ( $\downarrow$ 1.02)

Table 3: Results of ablation study for PQ-Net in average Macro-F1 on the test set of two benchmarks, where “w/o” means removal of a component from the original PQ-Net, and “ $\Delta$ ” represents the performance change.

average F1, respectively. Besides, on Pandora, PQ-Net also outperforms the two models by 2.30 and 2.66, respectively. These results demonstrate that the enhanced user representations by PQ-Net with psychological questionnaire knowledge are more effective for personality detection. Second, our PQ-Net is superior to AttRCNN which also introduces psychological features, showing that our model is more appealing in leveraging the prior psychological knowledge for personality detection. Third, the BERT-series models outperform the non-pretrained ones by considerable margins, showing the power of this pre-trained model in personality detection.

## 4 Analysis and Discussion

### 4.1 Ablation Study

The overall results above have demonstrated the effectiveness of our PQ-Net model as a whole. To further study the impact of each key module, we conduct an ablation study by removing them in turn from PQ-Net. The results, which are organized into three groups, are shown in Table 3. In the paragraphs below we only provide a detailed analysis on the Kaggle dataset, while similar conclusions can be obtained from the other dataset.

First, we investigate the contributions of the questions and choices in the questionnaire. When removing the question representation  $h^q$  in Eq. (4) and replacing the cross attention with saliency attention (Vu et al., 2020) without query, the performance declines by 2.14, demonstrating that the questions are helpful for retrieving personality-related cues. On the other hand, when removing the trait-specific potential answer representation

$h^{a(t)}$  in Eq. (9) and replacing it with  $z$  in Eq. (4), the performance declines by 2.69, showing that explicitly exploiting the user-agnostic choices in the questionnaire is more helpful than using only the information retrieved from the posts.

Second, we investigate the contributions of the soft gate and its supervision role. When replacing the soft gate with a hard gate, the performance declines by 2.34. This is most probably because the soft gate is smoother than the hard one. Besides, when removing  $J_q$  from Eq. (12), the performance declines by 2.58, showing that this extra supervision is worth considering for training PQ-Net.

Third, we investigate the contributions of the cross attention and two streams. When replacing the cross attention with cosine similarity, the performance declines by 1.59, suggesting that the attention mechanism is more capable of aligning the heterogeneous spaces of questionnaire and posts. Besides, when removing the questionnaire stream but keeping the additional supervision by directly using the post representations  $h^p$  in Eq. (1) to predict the choices  $g$  in Eq. (5), the performance drops by 2.91. When removing the context stream and using only the potential answer representation  $h^{a(t)}$  as the final representation  $u^{(t)}$  in Eq. (9), the performance drops by 1.28. This demonstrates that the design of the two-streams structure in PQ-Net is more favorable to capture the personality cues. Finally, after making the two streams share one BERT encoder, the performance only declines by 0.67, showing the staggering encoding capability of BERT.

### 4.2 Correlation Analysis

Intuitively, the performance of our model can be relevant to the accuracy of correctly “answering” the questions in the questionnaire. To qualitatively show this, we analyze the Macro-F1 scores of PQ-Net according to different numbers of correctly predicted questions. For this purpose, we first group the users in the Kaggle’s test set according to the numbers of correctly answered questions by PQ-Net, and then record the Macro-F1 scores in each group. The results in the four personality dimensions are plotted in Figure 3, where the x-axis represents the number of correctly predicted questions, and the y-axis is the Macro-F1 score. We can see that as the number of correctly predicted questions grows, the performance of PQ-Net increases accordingly. In particular, for all the personality traits, the performance of PQ-Net almost reaches 100%

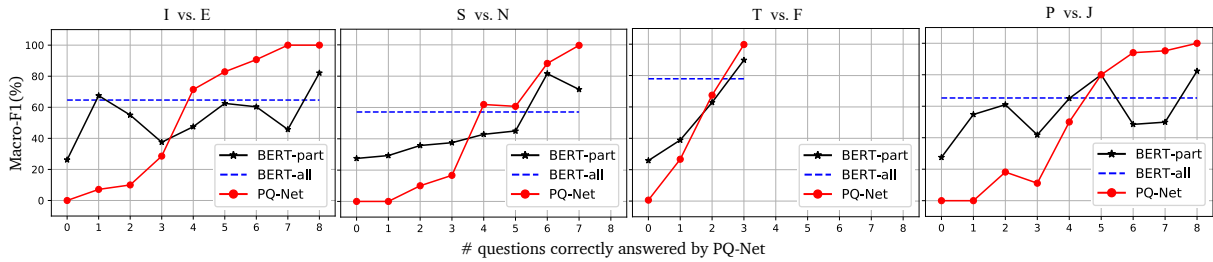


Figure 3: Performance of PQ-Net on Kaggle according to different numbers of correctly predicted questions.

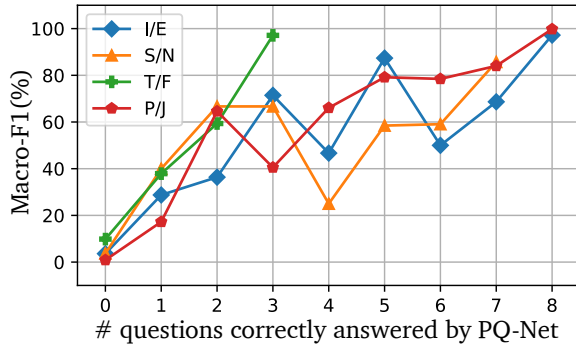


Figure 4: Results of correlation analysis on Pandora.

when all the questions are correctly predicted.

For comparison, we plot the results of two baselines for each group of users (BERT-part) and for all the users (BERT-all) respectively in Figure 3. One notable phenomenon is that PQ-Net suffers more from those groups with few questions predicted correctly than its counterparts. In the most extreme case, the performance of PQ-Net drops to 0% when all the questions are incorrectly answered. The reason behind this could be that PQ-Net overly relies on information from the questionnaire stream, so when the answers are incorrect, the information becomes misleading and deleterious.

Besides, we also plot the results of correlation analysis on the Pandora dataset in Figure 4, from which a similar trend can be observed as on Kaggle.

### 4.3 Visualization Analysis

The experimental results above have demonstrated the numerical performance of PQ-Net in detail. To further show whether the user representations enhanced by PQ-Net are inductive and distinguishable, we employ t-SNE (Laurens and Hinton, 2008) to reduce the dimension of the learned representations in the *T/F* trait of Kaggle’s test set to 2 and visualize the effect. As the results in Figure 5 show, PQ-Net visibly enforces a more compact clustering of examples in accordance with the personality labels than BERT, which benefits person-

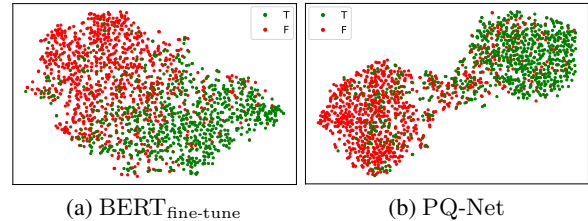


Figure 5: Visualization results of the produced user representations, in which (a) is the baseline without questionnaire enhanced, (b) is our PQ-Net.

ality classification accordingly. This experiment vividly demonstrates the superiority of our model in personality detection.

### 4.4 Case Study

In this subsection, we conduct the case study to further analyze our PQ-Net with a real example. As shown in Figure 6, we first record the probability of each choice of a question predicted by PQ-Net, and then plot the cross-attention weights from question to posts to show clues the model has discovered in the posts to support the decision. This experiment demonstrates that PQ-Net is able to judge the choices of question in the questionnaire via retrieving corresponding clues in the posts, providing an interpretability for personality detection.

## 5 Related Work

In recent years, numerous efforts have been devoted to automatically detecting one’s personality from his/her online texts (Adamopoulos et al., 2018; Tareaf et al., 2018; Guan et al., 2020). The early works rely on hand-crafted features (Yarkoni, 2010; Schwartz et al., 2013; Cui and Qi, 2017; Amirhosseini and Kazemian, 2020), which include various psycholinguistic features extracted by LIWC and statistical features extracted by bag-of-words models (Zhang et al., 2010). Nevertheless, feature engineering-based methods are limited by their ca-

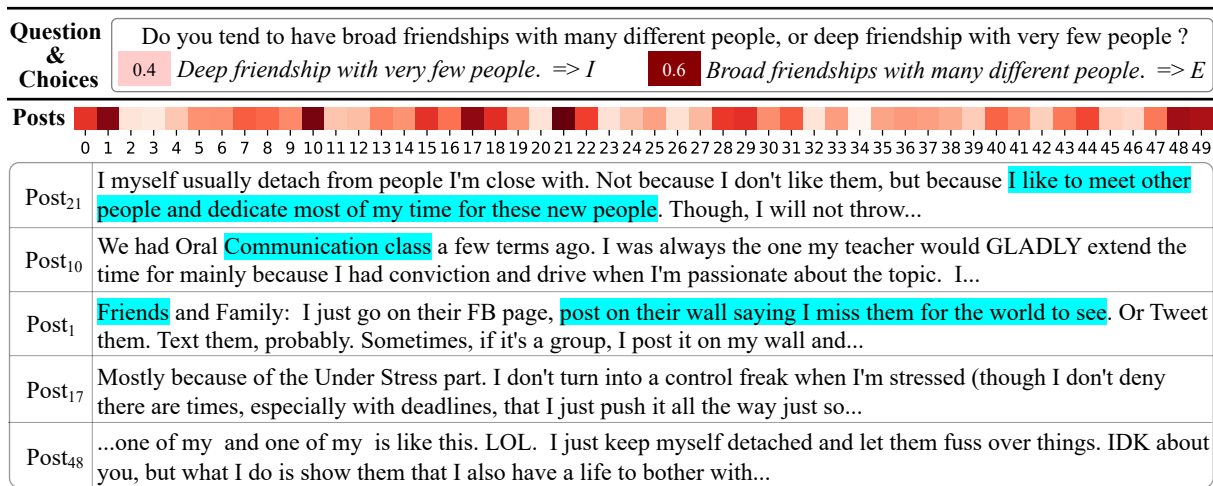


Figure 6: Results of case study in the *I/E* trait of Kaggle’s test set, where *E* is the ground truth personality label. At the top, the question asks about friendship with two choices corresponding to the Introversion and Extroversion personality, respectively. According to the prediction by PQ-Net, the probability of the user choosing “*Deep friendship with very few people*” and “*Broad friendships with many different people*” are 0.4 and 0.6, respectively. In the middle, the cross-attention weights from question to posts are shown. At the bottom, we show the contents of 5 posts with the highest attention weights and highlight the interpretable and related clues, e.g., “*I like to meet other people and dedicate most of my time for these new people*” in Post-21 is relevant to the questionnaire evidently.

pability in extracting many useful implicit features (Xue et al., 2018; Lynn et al., 2020).

Meanwhile, deep neural networks have been applied to personality detection by implicitly extracting features from the texts (Pradhan et al., 2020). For example, Hernandez and Knight (2017) and Tandra et al. (2017) applied LSTMs to encode each post with the GloVe embeddings (Pennington et al., 2014), and Keh et al. (2019) employed BERT to encode each post. Moreover, hierarchical structures were also applied to merge the texts into a user representation. For example, Lynn et al. (2020) first encoded each post via a gated recurrent unit (GRU) (Cho et al., 2014) with word attention, and then passed the encodings to a second GRU with post attention to aggregate the posts. Xue et al. (2018) designed an Inception (Szegedy et al., 2017) based AttRCNN module to encode each post and then applied a convolutional neural network (CNN) (Kalchbrenner et al., 2014) interact between posts. Despite numerous success, deep neural-network solutions rely on a data-driven fashion and lack the guidance of psychological domain knowledge.

## 6 Conclusion

In this paper, we proposed a psychological questionnaire enhanced network (PQ-Net) for personality detection. PQ-Net aims to track personality-related cues from online posts in an explicit manner

by considering the connections between the posts and a psychological questionnaire. Specifically, PQ-Net comprises a context stream and a questionnaire stream. The former encodes each post to obtain its contextual representation, and the latter learns to capture critical information in the posts to result in a potential answer representation for each question in the questionnaire. Finally, the potential answer representations are used to enhance the contextual post representations to predict the personality traits. Experimental results on two benchmarks show that PQ-Net outperforms the baselines significantly. Besides, further studies and analyses demonstrate that the representations enhanced by PQ-Net are more inductive and distinguishable, providing an interpretability for the personality detection process.

## 7 Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 62176270) and the Program for Guangdong Introducing Innovative and Entrepreneurial Teams (No. 2017ZT07X355).

## 8 Ethical Statement

This work aims to provide an innovative method to inspire research rather than creating a tool that violates privacy. The Kaggle and Pandora datasets used in our study are all anonymized, and we strictly abide by the use specifications of research



data and have not illegally detected any user privacy information. We realize that the evaluation results in this paper may have certain ethical risks and thus require that they should be used in a strictly controlled manner, subject to the approval of the institutional review board. Anyone who uses our work to secretly infer people’s personality characteristics is strictly prohibited.

## References

- Panagiotis Adamopoulos, Anindya Ghose, and Vilma Todri. 2018. The impact of user personality traits on word of mouth: Text-mining social media platforms. *Information Systems Research*, 29(3):612–640.
- Mohammad Hossein Amirhosseini and Hassan Kazemian. 2020. Machine learning approach to personality type prediction based on the myers–briggs type indicator®. *Multimodal Technologies and Interaction*, 4(1):9.
- Zan Mo Mo Aung and Phyu Hninn Myint. 2019. Personality prediction based on content of facebook users:a literature review. In *2019 20th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*.
- Katharine Cook Briggs and Isabel Briggs Myers. 1977. *The Myers-Briggs Type Indicator: Form G*. Consulting Psychologists Press.
- Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using rnn encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1724–1734.
- Paul T Costa Jr and Robert R McCrae. 1992. Four ways five factors are basic. *Personality and individual differences*, 13(6):653–665.
- Brandon Cui and Calvin Qi. 2017. Survey analysis of machine learning methods for natural language processing for mbti personality type prediction.
- P. S. Dandannavar, S. R. Mangalwede, and P. M. Kulkarni. 2020. Predicting the primary dominant personality trait of perceived leaders by mapping linguistic cues from social media data onto the big five model. In *International Conference on Advanced Machine Learning Technologies and Applications*.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186.
- Matej Gjurković, Mladen Karan, Iva Vukojević, Mihaela Bošnjak, and Jan Šnajder. 2020. Pandora talks: Personality and demographics on reddit. *arXiv preprint arXiv:2004.04460*.
- Matej Gjurković and Jan Šnajder. 2018. 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*, pages 87–97.
- Z. Guan, B. Wu, B. Wang, and H. Liu. 2020. **Personality2vec: Network representation learning for personality**. In *2020 IEEE Fifth International Conference on Data Science in Cyberspace (DSC)*, pages 30–37.
- R Hernandez and IS Knight. 2017. Predicting myers-bridge type indicator with text classification. In *Proceedings of the 31st Conference on Neural Information Processing Systems, Long Beach, CA, USA*, pages 4–9.
- Azhar Imran, Muhammad Faiyaz, and Faheem Akhtar. 2018. An enhanced approach for quantitative prediction of personality in facebook posts. *International Journal of Education and Management Engineering (IJEME)*, 8(2):8–19.
- Hang Jiang, Xianzhe Zhang, and Jinho D Choi. 2020. Automatic text-based personality recognition on monologues and multiparty dialogues using attentive networks and contextual embeddings (student abstract). In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 13821–13822.
- Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom. 2014. A convolutional neural network for modelling sentences. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 655–665.
- Sedrick Scott Keh, I Cheng, et al. 2019. Myers-briggs personality classification and personality-specific language generation using pre-trained language models. *arXiv preprint arXiv:1907.06333*.
- Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Van Der Maaten Laurens and Geoffrey Hinton. 2008. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(2605):2579–2605.
- Veronica Lynn, Niranjana Balasubramanian, and H Andrew Schwartz. 2020. Hierarchical modeling for user personality prediction: The role of message-level attention. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5306–5316.

- Matthew Matero, Akash Idnani, Youngseo Son, Salvatore Giorgi, Huy Vu, Mohammad Zamani, Parth Limbachiya, Sharath Chandra Guntuku, and H Andrew Schwartz. 2019. Suicide risk assessment with multi-level dual-context language and bert. In *Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology*, pages 39–44.
- Robert R McCrae and Paul T Costa Jr. 1989. Reinterpreting the myers-briggs type indicator from the perspective of the five-factor model of personality. *Journal of personality*, 57(1):17–40.
- Dong Nie, Zengda Guan, Bibo Hao, Shuotian Bai, and Tingshao Zhu. 2014. Predicting personality on social media with semi-supervised learning. In *2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, volume 2, pages 158–165. IEEE.
- Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. 2019. Pytorch: An imperative style, high-performance deep learning library. In *Advances in neural information processing systems*, pages 8026–8037.
- 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.
- Tejas Pradhan, Rashi Bhansali, Dimple Chandnani, and Aditya Pangaonkar. 2020. Analysis of personality traits using natural language processing and deep learning. In *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, pages 457–461. IEEE.
- H Andrew Schwartz, Johannes C Eichstaedt, Margaret L Kern, Lukasz Dziurzynski, Stephanie M Ramones, Megha Agrawal, Achal Shah, Michal Kosinski, David Stillwell, Martin EP Seligman, et al. 2013. Personality, gender, and age in the language of social media: The open-vocabulary approach. *PLoS one*, 8(9):e73791.
- Agnieszka Sorokowska, Anna Oleszkiewicz, Tomasz Frackowiak, Katarzyna Pisanski, Anna Chmiel, and Piotr Sorokowski. 2016. Selfies and personality: Who posts self-portrait photographs? *Personality and Individual Differences*, 90:119–123.
- Sanja Stajner and Seren Yenikent. 2020. [A survey of automatic personality detection from texts](#). In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 6284–6295, Barcelona, Spain (Online). International Committee on Computational Linguistics.
- Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander Alemi. 2017. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31.
- Michael M Tadesse, Hongfei Lin, Bo Xu, and Liang Yang. 2018a. Personality predictions based on user behavior on the facebook social media platform. *IEEE Access*, 6:61959–61969.
- Michael M. Tadesse, Hongfei Lin, Bo Xu, and Liang Yang. 2018b. Personality predictions based on user behaviour on the facebook social media platform. *IEEE Access*, 6:61959–61969.
- Tommy Tandra, Derwin Suhartono, Rini Wongso, Yen Lina Prasetyo, et al. 2017. Personality prediction system from facebook users. *Procedia computer science*, 116:604–611.
- Raad Bin Tareaf, Philipp Berger, Patrick Hennig, and Christoph Meinel. 2018. Personality exploration system for online social networks: Facebook brands as a use case. In *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, pages 301–309. IEEE.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008.
- Alessandro Vinciarelli and Gelareh Mohammadi. 2014. A survey of personality computing. *IEEE Transactions on Affective Computing*, 5(3):273–291.
- Thanh Vu, Dat Quoc Nguyen, and Anthony Nguyen. 2020. A label attention model for icd coding from clinical text. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*, pages 3335–3341. International Joint Conferences on Artificial Intelligence Organization. Main track.
- Di Xue, Lifa Wu, Zheng Hong, Shize Guo, Liang Gao, Zhiyong Wu, Xiaofeng Zhong, and Jianshan Sun. 2018. Deep learning-based personality recognition from text posts of online social networks. *Applied Intelligence*, 48(11):4232–4246.
- Feifan Yang, Xiaojun Quan, Yunyi Yang, and Jianxing Yu. 2021a. Multi-document transformer for personality detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 14221–14229.
- Hsin-Chang Yang and Zi-Rui Huang. 2019. Mining personality traits from social messages for game recommender systems. *Knowledge-Based Systems*, 165:157–168.
- Tao Yang, Feifan Yang, Haolan Ouyang, and Xiaojun Quan. 2021b. Psycholinguistic tripartite graph network for personality detection. *arXiv preprint arXiv:2106.04963*.

Tal Yarkoni. 2010. Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. *Journal of research in personality*, 44(3):363–373.

Yin Zhang, Rong Jin, and Zhi-Hua Zhou. 2010. Understanding bag-of-words model: a statistical framework. *International Journal of Machine Learning and Cybernetics*, 1(1-4):43–52.

Yinhe Zheng, Guanyi Chen, Minlie Huang, Song Liu, and Xuan Zhu. 2019. Personalized dialogue generation with diversified traits. *arXiv preprint arXiv:1901.09672*.

## **A Questionnaire**

The detailed content of the questionnaire adopted in our study is shown in Figure 7, in which the questions, including choices, are divided into four categories, each corresponding to one of the four traits of MBTI personality. Besides, we also mark the personality tendency represented by each choice.

### (a) Introversion or Extroversion

<p>Are you usually a good mixer with groups of people or rather quiet and reserved?</p> <input type="checkbox"/> Quiet and reserved. ⇒ Introversion <input type="checkbox"/> Good mixer with groups of people. ⇒ Extroversion	<p>In a large group, do you more often introduce others or get introduced?</p> <input type="checkbox"/> More often get introduced. ⇒ Introversion <input type="checkbox"/> More often introduce others. ⇒ Extroversion
<p>Among your friends, are you full of news about everybody, or one of the last to hear what is going on?</p> <input type="checkbox"/> One of the last to hear what is going on. ⇒ Introversion <input type="checkbox"/> Full of news about everybody. ⇒ Extroversion	<p>Do you talk easily to almost anyone for as long as you have to or find a lot to say only to certain people or under certain conditions?</p> <input type="checkbox"/> Find a lot to say only to certain people or under certain conditions. ⇒ Introversion <input type="checkbox"/> Talk easily to almost anyone for as long as you have to. ⇒ Extroversion
<p>Do you tend to have broad friendships with many different people, or deep friendship with very few people?</p> <input type="checkbox"/> Deep friendship with very few people. ⇒ Introversion <input type="checkbox"/> Broad friendships with many different people. ⇒ Extroversion	<p>Can the new people you meet tell what you are interested in right away or only after they really get to know you?</p> <input type="checkbox"/> Only after they really get to know you. ⇒ Introversion <input type="checkbox"/> Right away. ⇒ Extroversion
<p>When you are with the group of people, would you usually rather join in the talk of the group, or stand back and listen first?</p> <input type="checkbox"/> Stand back and listen first. ⇒ Introversion <input type="checkbox"/> Join in the talk of the group. ⇒ Extroversion	<p>Do you usually show your feelings freely or keep your feelings to yourself?</p> <input type="checkbox"/> Keep your feelings to yourself. ⇒ Introversion <input type="checkbox"/> Show your feelings freely. ⇒ Extroversion

### (b) Sensing or iNtuitioN

<p>If you were a teacher, would you rather teach facts-based courses or courses involving opinion or theory?</p> <input type="checkbox"/> Teach facts-based courses. ⇒ Sensing <input type="checkbox"/> Courses involving opinion or theory. ⇒ iNtuitioN	<p>Would you rather be considered a practical person, or An out-of-the-box-thinking person?</p> <input type="checkbox"/> A practical person. ⇒ Sensing <input type="checkbox"/> An out-of-the-box-thinking person. ⇒ iNtuitioN
<p>In doing something that many other people do would you rather invent a way of your own, or do it in the accepted way?</p> <input type="checkbox"/> Do it in the accepted way. ⇒ Sensing <input type="checkbox"/> Invent a way of your own. ⇒ iNtuitioN	<p>Would you rather have as a friend someone who is always coming up with new ideas or someone who has both feet on the ground?</p> <input type="checkbox"/> Someone who has both feet on the ground. ⇒ Sensing <input type="checkbox"/> Someone who is always coming up with new ideas. ⇒ iNtuitioN
<p>Do you usually get along better with realistic people or imaginative people?</p> <input type="checkbox"/> Realistic people. ⇒ Sensing <input type="checkbox"/> Imaginative people. ⇒ iNtuitioN	<p>Do you admire more the people who are normal-acting to never make themselves the center of attention or too original and individual to care whether they are the center of attention or not?</p> <input type="checkbox"/> normal-acting to never make themselves the center of attention. ⇒ Sensing <input type="checkbox"/> Individual to care whether they are the center of attention or not. ⇒ iNtuitioN
<p>In reading for pleasure, do you enjoy odd or original ways of saying things, or like writers to say exactly what they mean?</p> <input type="checkbox"/> Like writers to say exactly what they mean. ⇒ Sensing <input type="checkbox"/> Enjoy odd or original ways of saying things. ⇒ iNtuitioN	

### (c) Think or Feeling

<p>Do you more often let your heart rule your head or your head rule your heart?</p> <input type="checkbox"/> Your head rule your heart. ⇒ Thinking <input type="checkbox"/> Your heart rule your head. ⇒ Feeling	<p>Do you usually value emotion more than logic or value logic more than feelings?</p> <input type="checkbox"/> Value logic more than feelings. ⇒ Thinking <input type="checkbox"/> Value emotion more than logic. ⇒ Feeling
<p>Is it a higher compliment to be called a person of real feeling or a consistently reasonable person?</p> <input type="checkbox"/> A consistently reasonable person. ⇒ Thinking <input type="checkbox"/> A person of real feeling. ⇒ Feeling	

### (d) Perception or Judging

<p>When you go somewhere for the day, would you rather plan what you will do and when or just go!!</p> <input type="checkbox"/> Just go!! ⇒ Perception <input type="checkbox"/> Plan what you will do and When. ⇒ Judging	<p>Does following a schedule appeal to you, or cramp you?</p> <input type="checkbox"/> Cramp you. ⇒ Perception <input type="checkbox"/> Appeal to you. ⇒ Judging
<p>Does the idea of making a list of what you should get done over a weekend help you, or stress you, or positively depress you?</p> <input type="checkbox"/> Stress you or positively depress you. ⇒ Perception <input type="checkbox"/> Help you. ⇒ Judging	<p>When it is settled well in advance that you will do a certain thing at a certain time, do you find it nice to be able to plan accordingly or a little unpleasant to be tied down?</p> <input type="checkbox"/> A little unpleasant to be tied down. ⇒ Perception <input type="checkbox"/> Nice to be able to plan accordingly. ⇒ Judging
<p>When you have a special job to do, do you like to organize it carefully before you start or find out what is necessary as you go along?</p> <input type="checkbox"/> Find out what is necessary as you go along. ⇒ Perception <input type="checkbox"/> Organize it carefully before you start. ⇒ Judging	<p>Are you more successful at following a carefully worked out plan, or at dealing with the unexpected and seeing quickly what should be done.</p> <input type="checkbox"/> At dealing with the unexpected and seeing quickly what should be done. ⇒ Perception <input type="checkbox"/> At following a carefully worked out plan. ⇒ Judging
<p>Do you prefer to arrange picnics, parties etc, well in advance, or be free to do whatever to looks like fun when the time comes?</p> <input type="checkbox"/> Be free to do whatever to looks like fun when the time comes. ⇒ Perception <input type="checkbox"/> Stand back and listen first. ⇒ Judging	<p>In your daily work, do you.....</p> <input type="checkbox"/> Usually plan your work so you won't need to work under pressure and hate to work under pressure. ⇒ Perception <input type="checkbox"/> Rather enjoy an emergency that makes you work against time. ⇒ Judging

Figure 7: Detailed content of the questionnaire adopted in this study.