RESEARCH ARTICLE

# Using Psycholinguistic Clues to Index Deep Semantic Evidences: Personality Detection in Social Media Texts

**Qirui Tang[1], Wenkang Jiang[2], Xinlong Pan[3,*], Lei Lin[4], Jizhao Zhu[5], Yihua Du[4,*] and Donghong Sun[6]**

[1] School of Cyber Security, University of Chinese Academy of Sciences, Beijing 101408, China
[2] Australian Institute for Machine Learning, The University of Adelaide, Adelaide 5005, Australia
[3] Institute of Information Fusion, Naval Aviation University, Yantai 264001, China
[4] Computer Network Information Center, Chinese Academy of Sciences, Beijing 100085, China
[5] School of Computer Science, Shenyang Aerospace University, Shenyang 110136, China
[6] Institute for Network Science and Cyberspace, Tsinghua University, Beijing 100084, China

## Abstract

**Detecting personalities in social media content is an important application of personality psychology. Most early studies apply a coherent piece of writing to personality detection, but today, the challenge is to identify dominant personality traits from a series of short, noisy social media posts. To this end, recent studies have attempted to individually encode the deep semantics of posts, often using attention-based methods, and then relate them, or directly assemble them into graph structures. However, due to the inherently disjointed and noisy nature of social media content, constructing meaningful connections remains challenging. While such methods rely on well-defined relationships between posts, effectively capturing these connections in fragmented and sparse content is non-trivial, particularly under limited supervision or noisy input. To tackle this, we draw inspiration from the scanning reading technique—commonly recommended for efficiently processing large volumes of information—and propose an index attention mechanism as a solution. This mechanism leverages prior psycholinguistic knowledge as an "index" to guide attention, thereby enabling more effective information fusion across scattered semantic signals. Building on this idea, we introduce the Index Attention Network (IAN)—a novel framework designed to infer personality labels by performing targeted information fusion over deep semantic representations of individual posts. Through a series of experiments, IAN achieved state-of-the-art performance on the Kaggle dataset and performance comparable to graph convolutional networks (GCN) on the Pandora dataset. Notably, IAN delivered an average improvement of 13% in terms of macro-F1 scores with the Kaggle dataset. The code for IAN is available at GitHub: https://github.com/Once2gain/IAN.**

**Keywords**: personality detection, attention mechanism, social media text mining, information fusion.

**Academic Editor:**
Yulong Huang

**Submitted:** 04 March 2025
**Accepted:** 15 April 2025
**Published:** 26 April 2025

**Vol.** 2, **No.** 2, 2025.
10.62762/CJIF.2025.820998

***Corresponding authors:**
✉ Xinlong Pan
airadar@126.com
✉ Yihua Du
yhdu@cashq.ac.cn

**Citation**
Tang, Q., Jiang, W., Pan, X., Lin, L., Zhu, J., Du, Y., & Sun, D. (2025). Using Psycholinguistic Clues to Index Deep Semantic Evidences: Personality Detection in Social Media Texts. *Chinese Journal of Information Fusion*, 2(2), 112–126.

## 1 Introduction

Personality testing is a common task in psychology. The traditional method of testing a subject's personality is an artificially designed questionnaire, which is relatively reliable but not particularly efficient. To this end, researchers have proposed several different automated methods of testing personalities. Of these, analyzing user-generated content is one of the most important [1, 2]. Fortunately, social media provides vast quantities of user-generated content to test and, here, text, as the most abundant type of content, has proven to contain rich information that substantially reflects the author's individuality [3, 4]. It is this information that helps us to understand human cognitive and behavioral patterns. As such, personality detection has numerous promising applications in fields such as marketing and social network analysis. Figure 1 provides an outline of how personality detection works in social media texts.

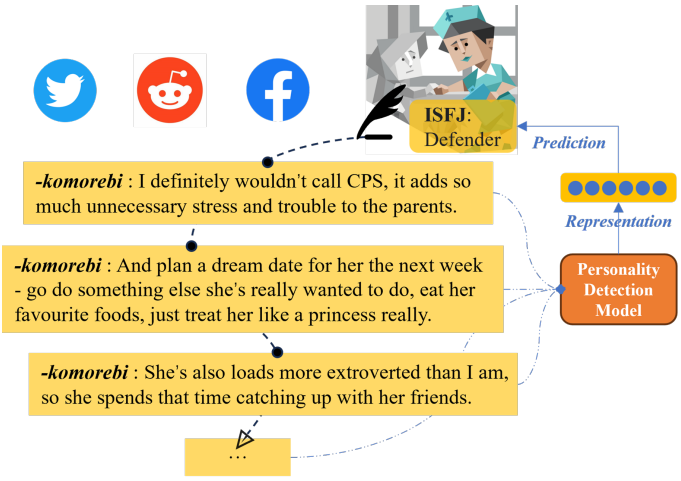


**Figure 1.** Personality detection from social media posts.

### 1.1 Language use and personality

Researchers have long been attempting to identify an individual's personality traits from their writings. Typically, they study the styles that surround people use words, with most articles reporting high within-person stability of language use, which has been linked to personality, psychological interventions, and other phenomena [4–7]. Moreover, this stability tends to persist no matter how the text is written: as a stream-of-consciousness, as an essay, or in self-narrative format [8–10].

That said, most of the above studies were conducted in a laboratory setting, where the subjects not only produced writing samples in the lab, but limitations were also placed on the topics to write about and the size of the sample. To address this issue, researchers have turned to more naturalistic writing texts, such as blogs on websites or posts on social media platforms [11–14]. Here, researchers have found inherent consistency in the personalities implied between the social media texts and normal, freely-written samples. For example, Gill et al. [39] conclude that bloggers tend to adapt to the possibilities of the medium rather than trying to present themselves differently.

### 1.2 Progression of Personality Detection in Texts

Researchers have also found that linguistic expressions have a significant non-linear correlation to personality traits [1, 6–9]. In fact, several of the initial advances made in personality detection with texts have been closely related to some of the research findings from psycholinguistics. For example, people who score high on extraversion generally use more social words, show more positive emotions, and tend to write more words but fewer large words [6, 15]. Linguistic Inquiry and Word Count (LIWC) [16] has been one of the most widely used tools for analyzing word use. Emotional dictionaries are also frequently mentioned [17, 18], as emotional experience has proven to be a key factor in personality analysis [19]. Further, many researchers are relying on statistical strategies to build a combined feature set, i.e., a set of word expressions, to feed into traditional machine learning models. The idea is to extract a linguistic feature pattern that can be used to predict personality traits [13, 14, 20] for a better result. These statistical features, such as LIWC, are frequently called (traditional) psycholinguistic features or psycholinguistic clues.

Additionally, numerous researchers have been working on feature engineering as a way of extracting personality-related signals from raw text [18, 21, 22]. For instance, Celli et al. [23] summarized two approaches to personality recognition: bottom-up and top-down. The bottom-up approach seeks cues from the data, like using n-gram features for text classification [11, 24]. Conversely, the top-down approaches use external resources, such as LIWC, to test correlations between word use and personality traits [3]. Further, with the rise of Transformers [25], researchers have turned to Transformer-based pre-trained language models (PLMs) to extract deep semantic features, also sometimes called PLM features. Such approaches have demonstrated encouraging progress [22, 26], generally relegating psycholinguistic features to second place.

One of the most recent advancements in personality

prediction has been to measure personality types from social media posts [26, 27]. Given the abundance of diverse text on social media, this has proven to be an easy-access approach that also saves time – especially compared to questionnaire-based approaches [28, 29]. Take blog posts as an example. These are typically short, topic-agnostic documents written in the author's natural style whose content may or may not contain personality clues [30]. To explore these posts for personality detection, one stream of research sees this content converted into graph networks. In turn, these networks reveal the inherent patterns in the structure of the posts [30–33].

In terms of prediction models, researchers have tried to integrate PLM features and psycholinguistic features using graph representations of the posts [32, 33]. Usually, the psycholinguistic features are used to construct the connection between posts, while the PLM features are used for the representations of each post.

Counter to this graph network-based approach, we propose an attention-based network called IAN. IAN weights the posts according to the psycholinguistic features and produces representations of the posts according to the PLM features, which are then subject to deep classification. Similar to the approaches based on graph networks, IAN uses the psycholinguistic features to mine correlations between posts, weighting each post according to the correlations found. As a more intuitive explanation, the psycholinguistic features are used to calculate the query or key for the self-attention mechanism, while the PLM features are used to calculate the self-attention value. In this way, IAN inherits all the advantages of the graph-based methods. That is, IAN uses the psycholinguistic features as clues to index evidence from the deep semantic features for personality prediction.

As a last point, one of the dilemmas faced by all existing studies is that, in situations where computing resources are limited, only a small portion of an author's posts can be taken as input. To remedy this issue, our IAN framework incorporates a topic clustering method. Thus, the contributions of this study are summarized as follows:

- We present an index attention mechanism for personality detection that can be thought of as a PLM-based multi-document classification mechanism. The basic principle is to use prior knowledge to pre-estimate the index score of each document against other documents so as to help aggregate task-specific features as distinct from pre-trained features.
- Experiments on two different datasets demonstrate that IAN is a highly effective approach to personality detection. Notably, its performance on the Kaggle dataset is, on average, 13% better than the current state-of-the-art performance in terms of Macro F1 scores.
- We visualized the index score matrices in the index attention mechanism as a way to summarize how the working patterns of index attention help facilitate information fusion across different segments. The results provide evidence that using psycholinguistic features to establish inter-document indexes can be quite effective.

## 2 Related Studies

### 2.1 Social Media Text-based Personality Detection

In order to conduct personality detection with social media posts, one has to jointly consider many short pieces of disparate text. This is an entirely different task from classifying long tracts of prose. Hence, recently, researchers have proposed several novel ways of enhancing the representation of social media posts by examining the interactions between them. Lynn et al. [34] point out that not all posts are equally important. Based on this idea, they proposed a hierarchical network based entirely on a GRU called SN+Attn. Within this framework, they use message-level attention to learn the weight of each post, while trying to recover high signal messages from noisy data. Similarly, Yang et al. [31] propose a post-order-agnostic encoder named Transformer-MD, which is a modified version of Transformer-XL. Their variant encodes any number of posts through memory tokens that represent previous posts. However, because noisy and scattered semantics are prone to interfere with each other, relying solely on self-attention to refine the representation of social media texts is not particularly wise [25].

The latest method to achieve optimal performance is the graph neural network (GNN), which has been used to model the structural relations between posts. In the graph, the nodes are posts, and the edges represent the similarity between the psycholinguistic features of the content. Generally, pre-trained language models are used to initialize the representations of posts as nodes. Yang et al. [32] tried to inject psycholinguistic knowledge into a heterogeneous graph called TrigNet by associating psycholinguistic category nodes to nodes of posts through word

nodes as intermediaries. Zhu et al. [33] proposed CGTN by constructing a second graph of social media posts initialized with psycholinguistic features and employing contrastive learning to determine graph similarity, while Yang et al. [30] developed D-DGCN through a dynamic multi-hop structure that automatically updates inter-post connections. Departing from these graph-based approaches, our study introduces a lightweight attention network enhanced with author-specific topic preference and psycholinguistic knowledge, eliminating the need for complex graph structures.

Recent advances in personality detection have seen a noticeable decline in attention-based and Transformer-based models. This shift is largely attributable to the inherent challenges Transformers face when applied to fragmented and short texts common in social media—specifically, their limited ability to capture user-level patterns from dispersed and heterogeneous linguistic signals. As a result, recent research has increasingly favored graph-based frameworks, which offer a natural means to model latent relationships between posts and facilitate joint learning of user trait representations. These models often incorporate psycholinguistic cues to establish meaningful inter-post connections, as illustrated by the works [32, 33]. While GCN-based methods have demonstrated effectiveness in modeling structural relationships between posts, they face limitations in both robustness and efficiency. These models depend heavily on the quality of graph construction, which can be problematic in the presence of fragmented or noisy user posts where meaningful edges are difficult to define. For instance, methods such as TrigNet and D-DGCN require complex mechanisms (e.g., heterogeneous nodes and dynamic multi-hop edge updates) to infer useful connections, leading to increased computational overhead and implementation complexity.

To address these limitations, we propose a lightweight attention-based framework that revisits the use of self-attention for modeling inter-post dynamics. Rather than relying on fixed graph structures, our method leverages learned attention weights guided by psycholinguistic priors and user-specific topic preferences, allowing for fine-grained control over post importance and interdependencies. In doing so, our approach maintains the relational modeling strength of graph-based methods while improving scalability and reducing computational complexity.

### 2.2 The Role of Psycholinguistic Knowledge

In previous studies, count-based psycholinguistic features are occasionally used as a supplement to deep learning features. Further, a common practice is to directly concatenate the vectors of two features [18, 22, 35]. However, in personality detection, this practice neglects the different natures and potential capabilities of the two types of features.

More specifically, deep semantic information has the ability to reflect an author's thoughts and feelings towards life, along with their behavioral characteristics. According to the American Psychological Association [1]: "personality refers to the enduring characteristics and behavior that comprise a person's unique adjustment to life, including major traits, interests, drives, values, self-concept, abilities, and emotional patterns". Pre-trained language models and deep neural networks help us to extract such patterns and link them to specific personality traits. (In this paper, we call them personality patterns.) However, a considerable number of posts consist solely of useless information [36], such as objective descriptions, which can hinder the detection of an author's personality.

By contrast, psycholinguistic features are too shallow to fully depict the personality of an author. But, fortunately, as prior knowledge, they do indicate where the text exposes personality. Notably, psycholinguistic features have found a resurgence in recent studies as a way of enriching the connections between posts. For example, these methods rely on psycholinguistic priors to establish edges between posts, thereby refining post representations within their graph learning frameworks [32, 33]. However, due to the high computational costs of processing so much text or the limited number of samples collected by existing datasets, most extant studies only use a random sample of posts when building graphs for each author. This obviously creates some limitations. First, the edges between posts could be false due to the limited selection of samples alongside the presence of noise. Second, training a GNN with too few posts can be difficult. Hence, in this study, we transformed the posts into semantic segments, which helps to increase the number of interactions between the texts. In addition, our approach is based on index attention, which efficiently harnesses the value of psycholinguistic knowledge when attempting to discover important semantics as evidence for personality detection.

[1]https://www.apa.org/topics/personality

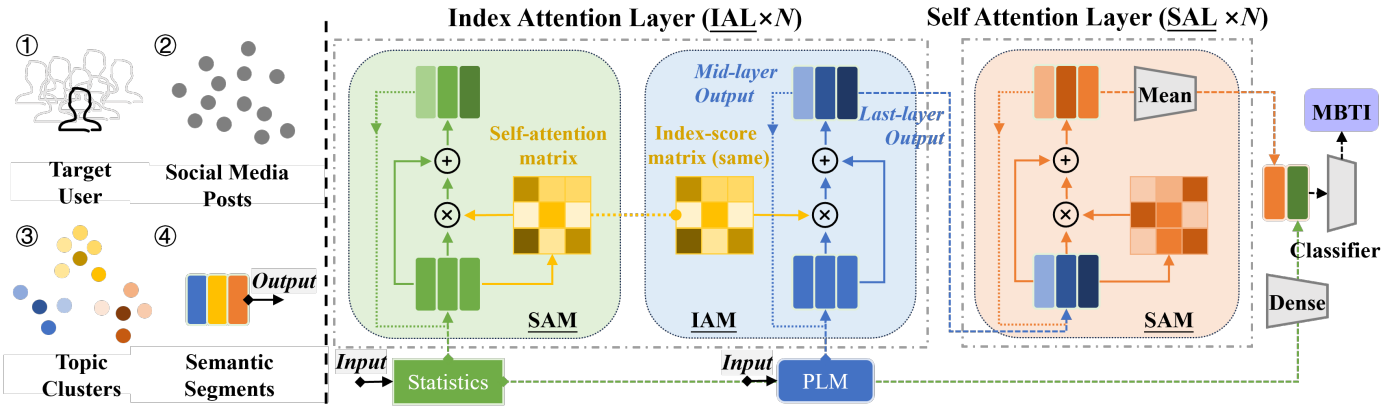


**Figure 2.** Overview of our methods. The left panel illustrates how social media posts are transformed into semantic segments, which serve as the inputs to the Index Attention Network (IAN). The three panels on the right depict the architecture of IAN, which consists of a psycholinguistic statistics tool, a pre-trained language model (PLM), N Index Attention Layers (IALs), N Self Attention Layers (SALs), and a classifier. Each IAL is composed of an Index Attention Mechanism and a Self Attention Mechanism, while each SAL consists solely of a Self Attention Mechanism.

## 3 Method

### 3.1 Overview

This section describes our proposed attention mechanism – Index Attention – which uses psycholinguistic features for the query and key calculations and deep semantic features for the value calculations. The approach was inspired by the scanning reading technique, where each post is scanned and weighted according to its psycholinguistic features, while the posts' deep semantic features are used to produce the representations. In addition, we recommend clustering the posts to screen for potential content that will help detect personalities. IAN is then built as a stack of index attention mechanisms for the purpose of summarizing the personality patterns scattered through the posts. Figure 2 depicts an overview of our methods.

### 3.2 Clustering and Sampling

In social media, an author may have created hundreds or even tens of thousands of posts. Yet, when taking computational efficiency into account, perhaps only a dozen or so can be sampled. Recent studies mostly consider 50-100 random posts, achieving relatively good detection performance from samples of this size [30, 31, 33]. On a different tangent, a series of studies have pointed out the correlations between a user's personality and their topic preferences [11, 37, 38], which inspired us to make a preliminary selection of posts by topics. For example, Gill et al. [39] find that bloggers who are high in 'openness' are likely to express their interests, opinions, and even feelings in the content of topics related to the arts and similar intellectual pursuits. Hence, we wondered whether leveraging a clustering strategy to screen out the topics that the authors are most interested in might lead to high-quality input for a detection model. This process, which is depicted on the left side of Figure 2, is described in more detail as follows:

• The first step is to produce semantic representations of the posts. For this, we use Sentence-BERT (SBERT), which is a modified variant of a pre-trained BERT model [40]. SBERT can generate semantically meaningful sentence embeddings that can then be compared using cosine-similarity or Euclidean distance. Moreover, these measures can be performed extremely efficiently, reducing the effort of clustering of 10,000 sentences from tens of hours down to mere seconds. Notably, accuracy does not suffer during this process.

• Next, topic clustering is performed based on the distance between posts. This is accomplished by HDBSCAN [41], which is a density-based clustering algorithm that can automatically determine the optimal number and shape of the resulting clusters.

• The last step is to draw samples from the largest clusters. Our framework considers the $n$ largest clusters, which represent the topics the author is most interested in and, therefore, most likely to talk about. The closest 5-10 posts to the center of each selected cluster are sampled and assembled into a segment of a few hundred words. Thus, the final sample consists of $n$ segments $S = \{s_1, s_2, \cdots, s_n\}$ of different topics, all with rich semantics.

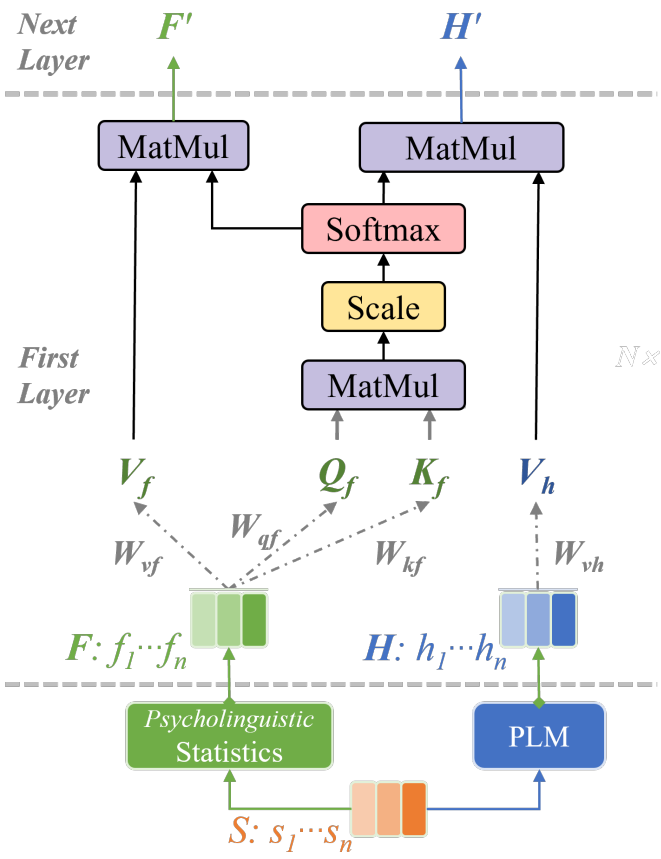


**Figure 3.** Index Attention Layer (IAL).

### 3.3 Index Attention Layer

As mentioned, we drew inspiration for the index attention mechanism from the scanning reading strategy, which is one of the speed reading techniques recommended by colleges and universities[2][3]. Scanning means you look only for specific pieces of information, such as a set of keywords, and once you locate a section requiring attention, you slow down and read it more thoroughly. In our attention mechanism, the psycholinguistic features are the specific pieces of information we are looking for, and the deep semantic features are the elements drawn from reading the posts more thoroughly. More specifically, we began with the self-attention mechanism outlined in [25], and modified it such that the query and key calculations are based on the psycholinguistic features, while the value calculations are based on the PLM features.

### 3.4 Index Attention Mechanism

As depicted in Figure 3, psycholinguistic features $F = \{f_1, f_2, \cdots, f_n\}, f_i \in F$ are extracted from $n$ segments $S = \{s_1, s_2, \cdots, s_n\}$, where $F$ is a vector of $p$ dimensions. PLM features $H = \{h_1, h_2, \cdots, h_n\}, h_i \in H$ are also extracted from these segments, where $H$ is a vector of $q$ dimensions. To implement the attention mechanism, four trainable parameter matrices are defined: $W_{qf}, W_{kf}, W_{vf}$, and $W_{vh}$. $W_{qf}, W_{kf}, W_{vf}$ have a dimension of $(p, p)$, whereas $W_{vh}$ has a dimension of $(q, q)$.

In the index attention mechanism, we first calculate the complete self-attention based on one type of feature, while calculating a sole Value based on another type of feature:

$$Q_f, K_f, V_f = FW_{qf}, FW_{kf}, FW_{vf} \tag{1}$$

$$V_h = HW_{vh} \tag{2}$$

where $Q_f$, $K_f$, $V_f$ are matrices with dimensions of $(n, p)$, and $V_h$ is a matrix with dimensions of $(n, q)$.

In the standard self-attention mechanism, the attention score matrix $M_{attn}$ is computed using the query $Q_f$ and key $K_f$ as follows:

$$M_{attn} = \text{Softmax}(Q_f K_f^T / \sqrt{p}) \tag{3}$$

where $M_{attn}$ is a vector of $n * n$ dimensions.

Specifically, the process begins with a matrix multiplication (**MatMul**) operation between $Q_f$ and $K_f$. This is followed by a scaling operation (**Scale**) to adjust the values, after which the softmax function (**Softmax**) is applied to normalize the scaled output, ensuring that the attention weights sum to one.

The resulting attention matrix $M_{attn}$ represents the potential associations between each feature $f_i$ and $f_j$ in the process of identifying personality patterns. This matrix is analogous to the edge values between post nodes introduced by Zhu et al. [33] in their psycholinguistic view graphs.

The core of the index attention mechanism involves using $M_{attn}$ as the index score matrix to apply on $V_h$ via a matrix multiplication (**MatMul**) operation to generate context (segment)-sensitive deep representations:

$$H' = M_{attn} V_h \tag{4}$$

In this study, each $h'$ in $H'$ is a hidden state encoding a type of personality pattern led by the corresponding segment $s_i$.

[2]https://www.student.unsw.edu.au/reading-strategies
[3]https://www.lc.cityu.edu.hk/ELSS/Resource/sas/index.htm

### 3.5 Stacking Index Attention

To enhance the model's expressiveness and higher-order relations, a series of index attention layers are stacked together. The output of each layer $H'$ becomes the PLM input for the next layer. Similarly, $F'$ in Equation 5 is the output of the self-attention mechanism's attention to the psycholinguistic features, which becomes the psycholinguistic input for the next layer.

$$F' = M_{attn} V_f \tag{5}$$

This process involves using $M_{attn}$ as the self-attention score matrix to apply on $V_f$ via a matrix multiplication (**MatMul**) operation.

### 3.6 Self Attention Layer

The self-attention layers in our framework are similar to the Transformer layers in the RoBERTa model [48], in that their purpose is to enhance the representation of each semantic segment from the perspective of purely semantic interaction. Each self-attention layer comprises a single self-attention mechanism—illustrated in Figure 2 and denoted as "SAL" in Figure 1—which follows the design proposed by Vaswani et al. [25].

To accomplish strengthening the context (segment)-related representation of the segment from a purely semantic perspective, $H'_{last}$ from the last index attention layer is updated to $H''$ according to the following:

$$Q, K, V = H'_{last} W'_q, H'_{last} W'_k, H'_{last} W'_v \tag{6}$$

$$H'' = \text{Softmax}(QK^T / \sqrt{q}) V \tag{7}$$

where $W'_q$, $W'_k$ and $W'_v$ are trainable parameter matrices with dimensions of $(q, q)$. This self-attention layer can also be stacked, and the representations of the semantic segments can be averaged to obtain a unique and robust representation $r$ of the blog's author:

$$r = Mean(H''_{last}) \tag{8}$$

where $H''_{last}$ refers to the output of the last self-attention layer.

### 3.7 The Objective Function

Following prior studies [18, 42], IAN predicts the label for a single dimension of an author's personality. To supplement the final personality prediction, an overall psycholinguistic feature vector, denoted as $\underline{f}$, is computed based on the content in $S$, similar to previous approaches [18, 35, 43]. The feature vector $\underline{f}$ is subsequently processed through a Dense layer to produce the output $r$, which is then concatenated with $r$. This concatenated representation serves as the input to the classifier, which outputs the logits for a binary prediction:

$$\text{Logit} = \text{Classifier}(r \oplus \text{Dense}(\underline{f})) \tag{9}$$

where $\oplus$ denotes the concatenation operation, and Logit contains the raw scores assigned by the classifier for each class.

Next, the Softmax function is applied to the logits to predict the probability of a particular personality class. Specifically, personality$(p) = 1$ indicates that the personality dimension $p$ is positively expressed, while personality$(p) = 0$ indicates a negative expression. The probability is calculated as follows:

$$\begin{aligned} &P(\text{personality}(p) = 1 | S, \sigma) \\ &= \frac{\exp(\text{logit}_{p=1})}{\exp(\text{logit}_{p=1}) + \exp(\text{logit}_{p=0})} \end{aligned} \tag{10}$$

where $S$ represents the input segments, and $\sigma$ denotes all the parameters of the IAN model except for those related to the calculated statistics in $\underline{f}$. Here, $\text{logit}_{p=1}$ refers to the raw score output by the classifier for the positive expression of the personality dimension $p$, whereas $\text{logit}_{p=0}$ refers to the raw score for the negative expression. The Softmax function normalizes these logits by exponentiating and dividing by the total sum, yielding a probabilistic distribution over the two possible personality outcomes. A higher probability for personality$(p) = 1$ suggests stronger positive expression for the dimension, and vice versa for personality$(p) = 0$.

Cross entropy loss is used to count the loss over $K$ samples:

$$\begin{aligned} \mathcal{L} &= -\sum_{k=1}^{K} y_k \cdot \log p_k + (1 - y_k) \cdot log(1 - p_k) \\ &= -\sum_{k=1}^{K} \log P(y_k | S, \sigma) \end{aligned} \tag{11}$$

where $y_k$ refers to the label of the $k$th sample, and $p_k$ refers to the predicted probability of the positive label of the $k$th sample. If some classes in the dataset are severely imbalanced, a weighted cross-entropy loss $\mathcal{L}_w$ is used to address the problem:

$$\mathcal{L}_w = w \cdot \mathcal{L} \tag{12}$$

where $w$ is the weight of the class corresponding to $y_k$ , usually set as the ratio of the number of samples in another class to the number in this class.

## 4 Experiments

### 4.1 Datasets

Following previous studies [30, 31], we validated our methods on two publicly available datasets, Kaggle and Pandora, and employed the Macro-F1 metric to measure performance. The two datasets were randomly divided into 6:2:2 for training, validation, and testing, respectively.

#### *4.1.1 Kaggle*

Kaggle is one of the most commonly used datasets for personality detection. It was collected from Twitter[4] by the Personality Caf'e forum. The dataset consists of 8675 groups of tweets and MBTI labels. Each group comprises the last 50 tweets posted by a user. This dataset is available at Kaggle[5].

#### *4.1.2 Pandora*

Pandora is a newly collected dataset from Reddit[6], assembled by TakeLab at the University of Zagreb [29]. It consists of 9067 groups of posts and MBTI labels. Each group ranges from dozens to tens of thousands of posts by an author. This dataset is available on TakeLab's website[7].

### 4.2 Psycholinguistic Statistics

We have comprehensively considered the composition of psycholinguistic features according to previous studies [18, 22, 33, 35], and built our **Statistic** tool for Mairesse Features [1], SenticNet Features [44, 45] and NRC Emotion Lexicon Features [46].

**Mairesse Features:** Mairesse Features comprise LIWC [16] features, MRC [47] features, and prosodic and utterance-type features. The LIWC dictionary annotates a word or a prefix with multiple categories involving parts of speech, emotions, society, and the environment, while the MRC psycholinguistic database provides unique annotations based on syntactic and psychological information.

**SenticNet Features:** SenticNet [44, 45] annotates a word with emotional values corresponding to introspection, temper, attitude, sensitivity, and polarity.

**NRC Emotion Lexicon Features:** The NRC Emotion Lexicon [46] annotates a word with the polarity values of anger, anticipation, disgust, fear, joy, negative, positive, sadness, surprise, trust, and charged – each on a scale of 11.

### 4.3 Baselines

We compared IAN with several of the highest performing baselines in recent studies. These included **SN+Attn** and **Transformer-MD** as attention-based methods and **TrigNet**, **CGTN**, and **D-DGCN** as graph-based methods. The details of these methods are described in *Related Studies*.

### 4.4 Implementation Details

**Network Architecture:** The proposed Index Attention Network (IAN) consists of a psycholinguistic statistical tool, a RoBERTa model, 2 stacking Index Attention Layers (IALs), 2 stacking Self Attention Layers (SALs), a Dense layer and a Classifier. Each IAL consists of a Self Attention Module (SAM) and a Index Attention Module (IAM). The psycholinguistic statistical tool processes all the words in each segment and generates a psycholinguistic feature vector with a dimensionality of 113, while the PLM features have a dimensionality of 768. Therefore, the input vector to SAM of IAL has a dimensionality of n × 113, where n is the number of segments, and the input vector to IAM of IAL has a dimensionality of n × 768. The Dense layer with input dimension 113 and output dimension 16 compresses the overall psycholinguistic feature vector based on all input posts of a user. The resulting 16-dimensional vector is concatenated with the 768-dimensional personality feature vector from the IAN module and passed to the Classifier that outputs a 2-dimensional logit. In addition, the architecture of the RoBERTa model follows the default configuration [48].

**Fine-tuning Settings:** We used the Fairseq [49] tool to implement our network and conduct all experiments. We chose RoBERTa as our pre-trained language model and downloaded its weights from HuggingFace[8]. For training, we set a composite initial learning rate of $1e^{-5}$ for only **RoBERTa** and $5e^{-5}$ for **the other components in the IAN**. A small learning rate is used to fine-tune RoBERTa in order to better adapt it to the domain-specific data and personality detection task. The psycholinguistic statistical tool is employed during input preprocessing but is not involved in model training. The IAN was optimized using the Adam optimizer with a weight decay of 0.1. We employed a linear learning rate scheduler with 300 warm-up updates. The batch size was set to 4, with an

[4]https://twitter.com
[5]https://www.kaggle.com/datasnaek/mbti-type
[6]https://www.reddit.com
[7]https://psy.takelab.fer.hr/datasets/all
[8]https://huggingface.co/roberta-base

update frequency of 8, which approximates the effect of a batch size of 32. We applied a dropout rate of 0.5 for the **Classifier** component. We trained **IAN** for 10 epochs, with mixed-precision (fp16) training enabled for computational efficiency.

**Clustering and Topic Sampling Strategy:** To capture semantic diversity while maintaining a fixed-size input for each user, we adopted different preprocessing strategies for the Pandora and Kaggle datasets based on the volume of user posts. In the Pandora dataset, where users often have thousands of posts, we applied topic modeling using BERTopic to each user's full post history. BERTopic, configured with HDBSCAN as its clustering backend, performs density-based clustering on UMAP-reduced Sentence-BERT embeddings and automatically infers the number of clusters, labeling outliers as noise (Topic = -1). We then selected up to 10 prominent clusters (excluding noise) and sampled up to 10 representative texts from each cluster, ranked by topic assignment probability. If fewer than 100 segments were collected, we filled the remainder by sampling from the most dominant topic to ensure uniform input size. In contrast, the Kaggle dataset contains significantly fewer posts per user (typically around 100). Instead of clustering, we randomly sampled 80 posts per user, following the overall median post count. These posts were then arranged in chronological order and grouped into 8 equally sized semantic segments.

For more details on the implementation and training procedure, please refer to the GitHub repository[9].

### 4.5 Overall Results

Tables 1 and 2 present the best results from our experiments for IAN alongside the best results published in the original papers for the baselines. The **first block** presents the performance of traditional and pre-trained baseline methods, while the **second block** shows the results of recent attention-based methods, and the **third block** contains recent graph-based methods. We attempted to stack IAN with more layers but found that detection performance reached its peak when N=3 for Kaggle and N=2 for Pandora. We reason that adding more layers confuses IAN as to which personality pattern representations to retain. Significantly, IAN (N=3) yielded a 13% lead over the existing best result with Kaggle. As for Pandora, IAN achieved comparable performance but is more lightweight and flexible.

[9]https://github.com/Once2gain/IAN

**Table 1.** Results of the models in terms of Macro-F1 (%) scores on Kaggle, where IAN (N=n) consists of n index attention layers and n self-attention layers.

| Methods | E/ I | S/ N | T/ F | J/ P | Avg |
|---|---|---|---|---|---|
| SVM | 53.34 | 47.75 | 76.72 | 63.03 | 60.21 |
| XGBoost | 56.67 | 52.85 | 75.42 | 65.94 | 62.72 |
| BiLSTM | 57.82 | 57.87 | 69.97 | 57.01 | 60.67 |
| BERT | 64.65 | 57.12 | 77.95 | 65.25 | 66.24 |
| SN+Attn | 65.43 | 62.15 | 78.05 | 63.92 | 67.39 |
| Transformer-MD | 66.08 | 69.10 | 79.19 | 67.50 | 70.47 |
| TrigNet | 69.54 | 67.17 | 79.06 | 67.69 | 70.86 |
| D-DGCN | 69.52 | 67.19 | 80.53 | 68.16 | 71.35 |
| CGTN | 71.12 | 70.44 | 80.22 | 72.64 | 73.61 |
| IAN (N=2) | 83.68 | 82.14 | 85.63 | 83.14 | 83.65 |
| IAN (N=3) | **87.92** | **84.48** | **87.16** | **86.89** | **86.61** |
| IAN (N=4) | 83.27 | 80.89 | 85.73 | 85.19 | 83.77 |

**Table 2.** Results of the models in terms of Macro-F1 (%) scores on Pandora, where "*cluster" means the posts were pre-processed with topic clustering.

| Methods | E/ I | S/ N | T/ F | J/ P | Avg |
|---|---|---|---|---|---|
| SVM | 44.74 | 46.92 | 64.62 | 56.32 | 53.15 |
| XGBoost | 45.99 | 48.93 | 63.51 | 55.55 | 53.50 |
| BiLSTM | 48.01 | 52.01 | 63.48 | 56.21 | 54.93 |
| BERT | 56.60 | 48.71 | 64.70 | 56.07 | 56.52 |
| SN+Attn | 54.60 | 49.19 | 61.82 | 53.64 | 54.81 |
| Transformer-MD | 55.26 | **58.77** | 69.26 | 60.90 | 61.05 |
| TrigNet | 56.69 | 55.57 | 66.38 | 57.27 | 58.98 |
| D-DGCN | 61.55 | 55.46 | **71.07** | **59.96** | 62.01 |
| IAN(N=2) | 57.85 | 55.23 | 64.61 | 57.84 | 58.88 |
| IAN (N=2,*cluster) | **62.67** | 58.33 | 69.34 | 59.31 | **62.41** |
| IAN (N=3,*cluster) | 57.24 | 56.32 | 63.00 | 57.63 | 58.55 |

### 4.6 Ablation Study

To investigate the contribution of sampling versus clustering the topics, we conducted an ablation study using the Pandora dataset, where we removed the clustering step. The Pandora dataset is fairly large, and having a pre-filtering and selecting step on posts would be quite beneficial. Normally, posts would be randomly selected to form segments. However, we see the clustering step does help improve the Macro-F1 (%) score by an average of 3.53%, as shown in Table 2.

## 5 Explanation of Index Attention Mechanism

To better understand how the Index Attention Mechanism (IAM) facilitates personality inference, we analyzed attention matrices at different layers and identified four primary attention patterns. Each pattern reveals a distinct way in which personality-related linguistic signals are aggregated

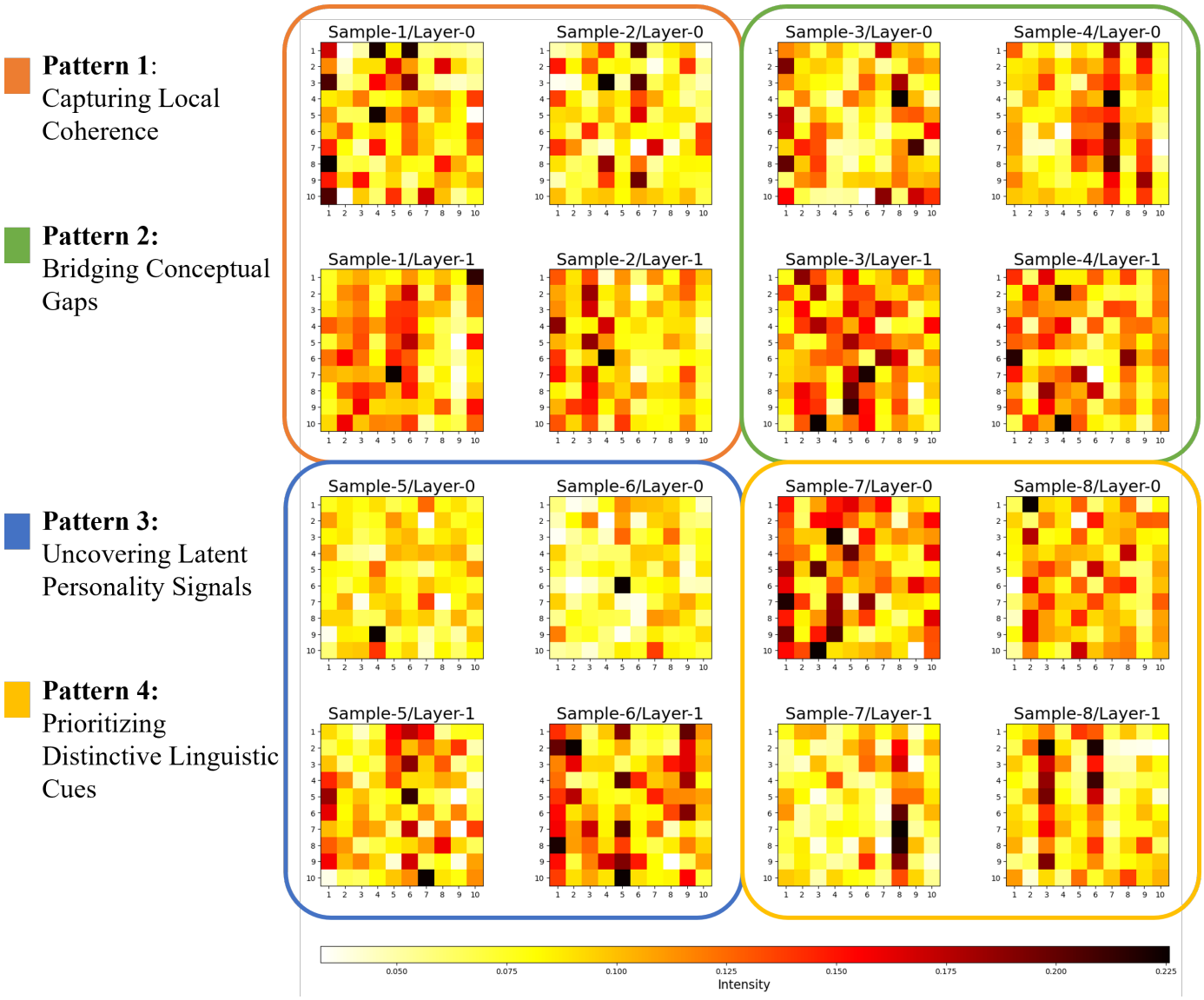


**Figure 4.** Index score matrices obtained from index attention layers at the inference stage.

across social media posts, which is crucial for addressing the challenge of fragmented and noisy data in personality detection.

These patterns were identified through experiments conducted with IAN (N=2, * cluster) trained on the Pandora dataset (see Table 2), where we visualized the typical index score matrices generated by the attention mechanism in Figure 4. The analysis of these matrices allows us to examine how IAM dynamically integrates psycholinguistic features to enhance personality trait recognition.

**Pattern 1: Capturing Local Coherence and Thematic Consistency (Visualized in Samples 1 and 2)**

At the first layer of IAM, the attention mechanism predominantly captures local proximity relationships between segments. This means that posts with apparent linguistic or topical similarities tend to receive higher mutual attention scores. However, when enriched with psycholinguistic features, IAM learns to generalize beyond surface similarities and establish deeper semantic connections.

This pattern is crucial for identifying personality traits that are manifested through consistent language use. For example, individuals with a Judging (J) preference in the MBTI framework often demonstrate structured and organized speech patterns across posts, reflecting their inclination toward planning and decisiveness. By detecting such proximities, IAM ensures that semantically related segments reinforce each other, improving personality classification accuracy.

**Pattern 2: Bridging Conceptual Gaps (Visualized in Samples 3 and 4)**

At the second layer, IAM shifts its focus to reinforcing relationships between segments that initially had weak or indirect associations. This suggests that the mechanism is effectively bridging gaps between posts that might not share direct lexical similarities but are conceptually related.

This pattern is particularly valuable in identifying personality traits that manifest through subtle contextual cues rather than explicit word choices. For example, individuals with an Extraversion (E) preference in the MBTI framework may use varied and context-dependent expressions of social engagement, such as discussing group activities, sharing spontaneous thoughts, or frequently interacting with others online. By learning these indirect associations, IAM improves its ability to detect latent personality signals dispersed across posts.

**Pattern 3: Uncovering Latent Personality Signals (Visualized in Samples 5 and 6)**

In this pattern, the initial correlations between segments based on surface-level psycholinguistic features appear weak. However, after processing through IAM, these correlations become significantly stronger. This suggests that IAM effectively integrates latent personality indicators that may not be immediately apparent in individual posts.

Many MBTI personality traits, such as Intuition (N), are characterized by abstract, exploratory, and metaphorical language rather than overtly repeated patterns. Intuitive individuals often discuss theoretical concepts, possibilities, and future-oriented ideas, which may appear loosely connected on the surface. This pattern demonstrates IAM's ability to extract deeper personality cues by leveraging psycholinguistic knowledge as an indexing mechanism, ensuring that even weakly correlated segments contribute meaningfully to the final inference.

**Pattern 4: Prioritizing Distinctive Linguistic Cues (Visualized in Samples 7 and 8)**

In cases where segments initially exhibit high similarity, IAM refines its focus by selectively amplifying key segments with more distinctive psycholinguistic features. These segments receive disproportionately high attention scores, suggesting that IAM prioritizes informative content over redundancy.

This pattern is particularly effective for differentiating between individuals with similar but distinct MBTI traits. For example, both Feeling (F) and Thinking (T) types may express strong opinions, but their linguistic focus differs: Feeling types tend to emphasize empathy, personal values, and emotional impact, whereas Thinking types prioritize logic, objectivity, and analytical reasoning. By prioritizing key segments, IAM ensures that personality inference is based on the most informative and distinguishing linguistic features rather than being skewed by redundant or less relevant content.

**Summary:** The identified attention patterns illustrate how IAM systematically refines personality-relevant linguistic signals across multiple layers. By leveraging prior psycholinguistic knowledge as an index, IAM successfully bridges weak connections, strengthens latent cues, and prioritizes informative content, thereby enhancing the robustness of personality detection in social media contexts.

## 6 Conclusion

The index attention mechanism presented in this paper leverages a set of prior (psycholinguistic) features to facilitate task-specific information fusion across documents—capturing correlations that pre-trained language models may overlook. Centered around each document, this mechanism enables the framework to integrate relevant information from other documents, thereby enhancing the robustness of task-specific representations.

Our implementation is specifically designed for multi-document classification tasks. It uses prior features to predict task-specific relationships between documents and performs information fusion by aggregating effective signals from deep semantics to refine each document's representation. In this study, we applied index attention to exploit the full potential of psycholinguistic knowledge as a clue for indexing and fusing evidence for personality detection from PLM features. We also developed an Index Attention Network (IAN) to detect personality traits from social media posts. IAN seeks to uncover deep semantic evidence through topic preferences, semantic relevance, and psycholinguistic cues. Experimental results on two publicly available datasets demonstrate the effectiveness of our methods.

## 7 Limitations and Ethical Considerations

This study has several noteworthy limitations. First, although our index attention mechanism takes

psycholinguistic features as prior knowledge, these features are fixed and do not have the same adaptability as deep neural network features, which can be pre-trained and fine-tuned for downstream tasks. Consequently, psycholinguistic features could be replaced by other feature engineering models, such as a CNN, pre-trained or not, to potentially extract better features than psycholinguistic ones for attention.

The second limitation concerns long text classification tasks. This method is more suitable for text fragments with sparse correlations and less appropriate for tightly contextual, long text classification. The principle of index attention is to filter out irrelevant posts, which may restrict its applicability in certain scenarios.

The third limitation relates to preprocessing. We sampled 100 posts per topic for each blog author, which evidently missed many posts relevant to personality assessment. Extending the index attention mechanism to integrate all sampled posts could enhance the precision of personality recognition.

Future research directions could explore ways to overcome these limitations, such as developing trainable feature extraction methods, improving the ability to handle tightly contextual information, and optimizing sampling strategies to enhance the overall model performance.

The use of computational methods, particularly machine learning and deep learning algorithms, for MBTI personality assessment based on online data raises several ethical concerns. While these advanced techniques offer improved applicability compared to traditional psychological scales, they present challenges in terms of interpretability and ethical implications. The "black box" nature of deep learning algorithms lacks a solid grounding in psychological theory, making their outcomes difficult to explain and potentially hindering the understanding of personality traits within these models. Furthermore, the vast amounts of data required for these algorithms raise significant privacy concerns. The Pandora dataset, for example, which adheres to ethical codes for psychological research, does allow for the use of archival data without individual consent – but only under specific conditions that protect participants from risk and where confidentiality is maintained [29]. Likewise, the Kaggle dataset has been anonymized to protect privacy.

## Data Availability Statement

The source code used in this study is publicly available on GitHub at the following link: https://github.com/Once2gain/IAN.

## Funding

This work was supported by the Natural Science Foundation of Shandong Province under Grant ZR2020MF154.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Ethical Approval and Consent to Participate

This study utilizes an anonymized public dataset, which is publicly available and does not contain any personally identifiable information. As the dataset is fully anonymized and used without the collection of personal data from individuals, ethical approval is not required for this research.

## References

[1] Mairesse, F., Walker, M. A., Mehl, M. R., & Moore, R. K. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of artificial intelligence research, 30*, 457-500. [CrossRef]

[2] Holtgraves, T. (2011). Text messaging, personality, and the social context. *Journal of research in personality, 45*(1), 92-99. [CrossRef]

[3] Lee, C. H., Kim, K., Seo, Y. S., & Chung, C. K. (2007). The relations between personality and language use. *The Journal of general psychology, 134*(4), 405-413. [CrossRef]

[4] Fast, L. A., & Funder, D. C. (2008). Personality as manifest in word use: Correlations with self-report, acquaintance report, and behavior. *Journal of personality and social psychology, 94*(2), 334.

[5] Schnurr, P. P., Rosenberg, S. D., Oxman, T. E., & Tucker, G. J. (1986). A methodological note on content analysis: Estimates of reliability. *Journal of personality assessment, 50*(4), 601-609. [CrossRef]

[6] Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: language use as an individual difference. *Journal of personality and social psychology, 77*(6), 1296.

[7] Pennebaker, J. W., Mehl, M. R., & Niederhoffer, K. G. (2003). Psychological aspects of natural language use: Our words, our selves. *Annual review of psychology, 54*(1), 547-577. [CrossRef]

[8] Mehl, M. R., Gosling, S. D., & Pennebaker, J. W. (2006). Personality in its natural habitat: manifestations and implicit folk theories of personality in daily life. *Journal of personality and social psychology, 90*(5), 862.

[9] Hirsh, J. B., & Peterson, J. B. (2009). Personality and language use in self-narratives. *Journal of research in personality, 43*(3), 524-527. [CrossRef]

[10] Ireland, M. E., & Pennebaker, J. W. (2010). Language style matching in writing: synchrony in essays, correspondence, and poetry. *Journal of personality and social psychology, 99*(3), 549.

[11] Nowson, S., & Oberlander, J. (2007, March). Identifying more bloggers: Towards large scale personality classification of personal weblogs. In *Proceedings of the international conference on weblogs and social*.

[12] Yarkoni, T. (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. [CrossRef]

[13] Golbeck, J., Robles, C., & Turner, K. (2011). Predicting personality with social media. In *CHI'11 extended abstracts on human factors in computing systems* (pp. 253-262). [CrossRef]

[14] Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Dziurzynski, L., Ramones, S. M., Agrawal, M., ... & Ungar, L. H. (2013). Personality, gender, and age in the language of social media: The open-vocabulary approach. *PloS one, 8*(9), e73791. [CrossRef]

[15] Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology, 29*(1), 24-54. [CrossRef]

[16] Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates, 71*(2001), 2001.

[17] Poria, S., Gelbukh, A., Agarwal, B., Cambria, E., & Howard, N. (2013). Common sense knowledge based personality recognition from text. In *Advances in Soft Computing and Its Applications: 12th Mexican International Conference on Artificial Intelligence, MICAI 2013, Mexico City, Mexico, November 24-30, 2013, Proceedings, Part II 12* (pp. 484-496). Springer Berlin Heidelberg. [CrossRef]

[18] Majumder, N., Poria, S., Gelbukh, A., & Cambria, E. (2017). Deep learning-based document modeling for personality detection from text. *IEEE intelligent systems, 32*(2), 74-79. [CrossRef]

[19] Watson, D., & Clark, L. A. (1992). On traits and temperament: General and specific factors of emotional experience and their relation to the five-factor model. *Journal of personality, 60*(2), 441-476. [CrossRef]

[20] Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., ... & Seligman, M. E. (2015). Automatic personality assessment through social media language. *Journal of personality and social psychology, 108*(6), 934.

[21] Sun, X., Liu, B., Cao, J., Luo, J., & Shen, X. (2018, May). Who am I? Personality detection based on deep learning for texts. In *2018 IEEE international conference on communications (ICC)* (pp. 1-6). IEEE. [CrossRef]

[22] Mehta, Y., Fatehi, S., Kazameini, A., Stachl, C., Cambria, E., & Eetemadi, S. (2020, November). Bottom-up and top-down: Predicting personality with psycholinguistic and language model features. In *2020 IEEE international conference on data mining (ICDM)* (pp. 1184-1189). IEEE. [CrossRef]

[23] Celli, F., Pianesi, F., Stillwell, D., & Kosinski, M. (2013). Workshop on computational personality recognition: Shared task. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 7, No. 2, pp. 2-5). [CrossRef]

[24] Oberlander, J., & Nowson, S. (2006, July). Whose thumb is it anyway? Classifying author personality from weblog text. In *Proceedings of the COLING/ACL 2006 main conference poster sessions* (pp. 627-634).

[25] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems, 30*.

[26] Christian, H., Suhartono, D., Chowanda, A., & Zamli, K. Z. (2021). Text based personality prediction from multiple social media data sources using pre-trained language model and model averaging. *Journal of Big Data, 8*(1), 68. [CrossRef]

[27] Han, S., Huang, H., & Tang, Y. (2020). Knowledge of words: An interpretable approach for personality recognition from social media. *Knowledge-Based Systems, 194*, 105550. [CrossRef]

[28] Gjurković, M., & Šnajder, J. (2018, June). 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* (pp. 87-97). [CrossRef]

[29] Gjurković, M., Karan, M., Vukojević, I., Bošnjak, M., & Šnajder, J. (2020). PANDORA talks: Personality and demographics on Reddit. *arXiv preprint arXiv:2004.04460*.

[30] Yang, T., Deng, J., Quan, X., & Wang, Q. (2023, June). Orders are unwanted: dynamic deep graph convolutional network for personality detection. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 37, No. 11, pp. 13896-13904). [CrossRef]

[31] Yang, F., Quan, X., Yang, Y., & Yu, J. (2021, May). Multi-document transformer for personality detection. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 35, No. 16, pp. 14221-14229). [CrossRef]

[32] Yang, T., Yang, F., Ouyang, H., & Quan, X. (2021). Psycholinguistic tripartite graph network for personality detection. *arXiv preprint arXiv:2106.04963*.

[33] Zhu, Y., Hu, L., Ge, X., Peng, W., & Wu, B. (2022). Contrastive Graph Transformer Network for Personality Detection. In *IJCAI* (pp. 4559-4565).

[34] Lynn, V., Balasubramanian, N., & Schwartz, H. A. (2020, July). 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* (pp. 5306-5316). [CrossRef]

[35] Ren, Z., Shen, Q., Diao, X., & Xu, H. (2021). A sentiment-aware deep learning approach for personality detection from text. *Information Processing & Management, 58*(3), 102532. [CrossRef]

[36] Štajner, S., & Yenikent, S. (2021, April). Why is MBTI personality detection from texts a difficult task?. In *Proceedings of the 16th conference of the European chapter of the association for computational linguistics: main volume* (pp. 3580-3589). [CrossRef]

[37] Liu, Y., Wang, J., & Jiang, Y. (2016). PT-LDA: A latent variable model to predict personality traits of social network users. *Neurocomputing, 210*, 155-163. [CrossRef]

[38] Zhao, J., Zeng, D., Xiao, Y., Che, L., & Wang, M. (2020). User personality prediction based on topic preference and sentiment analysis using LSTM model. *Pattern Recognition Letters, 138*, 397-402. [CrossRef]

[39] Gill, A., Nowson, S., & Oberlander, J. (2009, March). What are they blogging about? Personality, topic and motivation in blogs. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 3, No. 1, pp. 18-25). [CrossRef]

[40] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019, June). 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)* (pp. 4171-4186). [CrossRef]

[41] Campello, R. J., Moulavi, D., & Sander, J. (2013, April). Density-based clustering based on hierarchical density estimates. In *Pacific-Asia conference on knowledge discovery and data mining* (pp. 160-172). Berlin, Heidelberg: Springer Berlin Heidelberg. [CrossRef]

[42] El-Demerdash, K., El-Khoribi, R. A., Shoman, M. A. I., & Abdou, S. (2022). Deep learning based fusion strategies for personality prediction. *Egyptian Informatics Journal, 23*(1), 47-53. [CrossRef]

[43] KN, P. K., & Gavrilova, M. L. (2021). Latent personality traits assessment from social network activity using contextual language embedding. *IEEE Transactions on Computational Social Systems, 9*(2), 638-649. [CrossRef]

[44] Cambria, E., Liu, Q., Decherchi, S., Xing, F., & Kwok, K. (2022). SenticNet 7: A commonsense-based neurosymbolic AI framework for explainable sentiment analysis. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference* (pp. 3829–3839).

[45] Susanto, Y., Livingstone, A. G., Ng, B. C., & Cambria, E. (2020). The hourglass model revisited. *IEEE Intelligent Systems, 35*(5), 96-102. [CrossRef]

[46] Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word–emotion association lexicon. *Computational intelligence, 29*(3), 436-465. [CrossRef]

[47] Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology Section A, 33*(4), 497-505. [CrossRef]

[48] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

[49] Ott, M., Edunov, S., Baevski, A., Fan, A., Gross, S., Ng, N., ... & Auli, M. (2019). fairseq: A fast, extensible toolkit for sequence modeling. *arXiv preprint arXiv:1904.01038*.

**Qirui Tang** received the B.S. degree in Software Engineering from Communication University of China, Beijing, China, in 2021. He received the M.S. degree in Computer Technology at the University of Chinese Academy of Sciences, Beijing, China, in 2024. He is currently working toward the Ph.D. degree in Cyberspace Security at the University of Chinese Academy of Sciences, Beijing, China. (Email: tangqirui21@mails.ucas.ac.cn)

**Wenkang Jiang** received the B.S. degree in Computer Science and Technology from Hefei University of Technology, Hefei, China, in 2021. He received the M.S. degree in Computer Technology at the University of Chinese Academy of Sciences, Beijing, China, in 2024. He is currently working toward the Ph.D. degree in Computer Science with Australian Institute for Machine Learning, the University of Adelaide, Adelaide, Australia. (Email: wenkang.jiang@adelaide.edu.au)

**Xinlong Pan** received his B.Sc. and M.Sc. degrees in radar engineering in 2006 and 2010 respectively from the Air Force Radar Academy, China. He received the Ph.D. degree in information and communication engineering in 2017 from Naval Aviation University, China. Now he serves as an associate professor in Naval Aviation University. His research interests include geographic information engineering, trajectory data mining, intelligent information processing, and expert systems. (Email: airadar@126.com)

**Lei Lin** received the B.S. in applied mathematics from Northeast normal university, China, in 2009. He received the M.S. degrees in cryptography from Beijing University of Posts and Telecommunications, China, in 2012. And he got the Ph.D. degree in industrial engineering from Strasbourg University, Strasbourg, France, in 2016. From 2018 to 2022, he was a Post Doctor in the Institute of Computing Technology, Chinese Academy of Sciences. Currently, he is assistant researcher in the Computer Network Information Center, University of Chinese Academy of Sciences. His research interests include social computing, data mining, and innovation management. (Email: linlei@cnic.cn)

**Jizhao Zhu** received the Ph.D. degree in computer application technology from the Northeastern University, in 2018. He is currently a Master's Supervisor with the School of Computer, Shenyang Aerospace University. His research interests include machine learning, knowledge graph, and natural language understanding. (Email: zhujz@sau.edu.cn)

**Yihua Du** is currently a professorate senior engineer in Computer Network Information Center, Chinese Academy of Sciences. He has been engaged in management system research and development, data analysis and processing, network publishing platform construction, new media communication research and other work for many years. His current research interests include communication analysis and guidance, and software engineering techniques.(Email: yhdu@cashq.ac.cn)

**Donghong Sun** is an Associate Professor at Tsinghua University. Her research interest lies in Cyberspace security, including network detection and security assessment privacy computing, etc. (Email: sundonghong@tsinghua.edu.cn)