Knowledge-Based Systems 249 (2022) 108952

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

A lexical psycholinguistic knowledge-guided graph neural network for interpretable personality detection

Yangfu Zhu [a], Linmei Hu [a], Nianwen Ning [b], Wei Zhang [a], Bin Wu [a,*]

[a] Beijing Key Laboratory of Intelligence Telecommunication Software and Multimedia, Beijing University of Posts and Telecommunications, Beijing 100876, PR China
[b] School of Artificial Intelligence, Henan University, Zhengzhou 450046, PR China

ARTICLE INFO

Article history:
Received 14 May 2021
Received in revised form 8 April 2022
Accepted 28 April 2022
Available online 6 May 2022

Keywords:
Personality detection
Psycholinguistic knowledge
Heterogeneous network
Graph neural network

ABSTRACT

With the blossoming of online social media, personality detection based on user-generated content has a significant impact on information scientific and industrial applications. Most existing approaches rely heavily on semantic features or superficial psycholinguistic statistical features calculated by existing tools and fail to effectively exploit psycholinguistic knowledge that can help determine and interpret peoples personality traits. In this paper, we propose a novel lexical psycholinguistic knowledge-guided graph neural model for interpretable personality detection, which leverages the personality lexicons as a bridge for injecting relevant external knowledge to enrich the semantics of a document. Specifically, we learn a kind of personality-aware word embedding, that encodes psycholinguistic information in the continuous representations of words. Then, a Heterogeneous Personality word graph is constructed by aligning the personality lexicons with the personality knowledge graph, which is fed into a Message-passing graph Network (HPMN) to extract explicit lexicon and knowledge relations through the interactions among heterogeneous graph nodes. Finally, through a carefully designed readout function, all heterogeneous nodes are selectively incorporated as knowledge-guided document embeddings for user-generated text personality understanding and interpretation. Experiments show that our model effectively detects personality traits. Moreover, it provides a certain level of support for lexical hypotheses in psycholinguistic research from a computational linguistics perspective.

1. Introduction

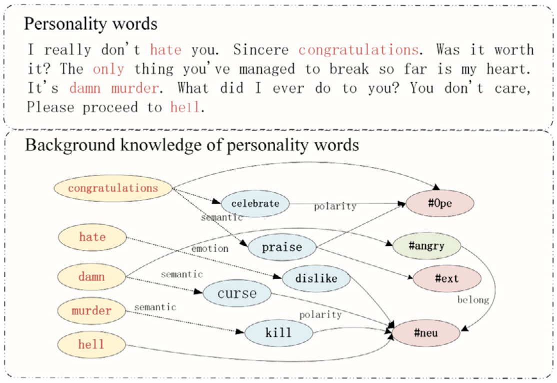
With the rapid development of social media platforms, people can access and analyze much user-generated content (UGC) to automatically identify authors personality traits. Many studies have shown that automatic personality detection systems play an essential role in various applications, such as user interest mining [1], information dissemination [2], recommendation systems [3–5], and intelligent machine design [6]. Therefore, analyzing and detecting users' personality traits is significant for grasping users' current and future psychologies and predicting their reactions and behaviors.

Personality detection research based on user-generated text is mainly divided into psycholinguistic lexicon-based, neural language model-based, and interpretability research. Earlier researchers captured psycholinguistic lexicon statistics features such as Linguistic Inquiry and Word Count (LIWC) [7] and Medical Research Council (MRC) [8] features in texts for personality detection [9,10]. However, the obtaining artificial features are a costly operation, and a statistical analysis cannot effectively represent the original semantics. To avoid feature engineering, deep neural models are employed to learn text-distributed representations from end to end, and the resulting detection accuracy is greatly improved [11–13]. However, neural language model embeddings lack the ability to explain personality. Recently, some researchers combined common knowledge to detect personality [14,15], providing some ability to explain personality and contributing to the analysis of personality traits. The latest researchers employed interpretable machine learning to clearly quantify the impacts of various psycholinguistic statistical features [16,17]. However, these methods do not deeply exploit psycholinguistic domain knowledge and fail to effectively integrate psycholinguistic knowledge and text semantics into the associated neural models.

In the psychology field, personality traits are defined as attribute combinations of individual thoughts and emotions to explain the differences in human behaviors [18]. The generally used measurement metric are the Big Five personality, including openness, conscientiousness, extroversion, agreeableness, and neuroticism [19]. The relationship between personality and language has been studied for a long time. Psycholinguistics found

[*] Corresponding author.
E-mail address: wubin@bupt.edu.cn (B. Wu).

https://doi.org/10.1016/j.knosys.2022.108952
0950-7051/© 2022 Published by Elsevier B.V.


Personality words

I really don't hate you. Sincere congratulations. Was it worth it? The only thing you've managed to break so far is my heart. It's damn murder. What did I ever do to you? You don't care, Please proceed to hell.

Background knowledge of personality words

Fig. 1. An example of a neurotic user.

an interesting phenomenon in empirical research: personality traits affect people's use of language, which refers to their choice of vocabulary. Specifically, the LIWC lexicon [20,21] and some personality adjectives [22] (Personality Adjectives Check List)[1] have linear correlations with each personality trait. In addition, people with the same personality traits usually have the same fixed emotional polarities [23]. The details regarding this topic are described in Appendix. Fig. 1 shows a visual example of a neurotic user's psycholinguistic knowledge. From the words "*hate*", "*murder*", and "*hell*", we can roughly infer that he/she is a neurotic user. Based on the relationship between the synonym "*damn*", emotional polarity, and personality traits, this inference is more confident to be confirmed. It can be seen that conducting personality detection research from the lexical psycholinguistic knowledge perspective can bring rich domain structure knowledge rather than superficial psycholinguistic statistical information. Although research on personality detection has achieved remarkable results, some challenges still remain.

- Fusion of text semantics and psycholinguistic knowledge: It is a challenge to fully fuse lexical psycholinguistic knowledge and text semantics while accurately representing the personality traits derived from the user's language.
- Interpretability of personality detection: It is a challenge to utilize personality psychology knowledge to realize explainable personality detection in neural models.

To meet the above challenges, we propose a novel lexical psycholinguistic knowledge-guided graph neural network model for interpretable personality detection. Our model enriches personality document representations by incorporating heterogeneous external knowledge through the use of personality lexicons as intermediaries. In particular, instead of directly using previous pretrained word embeddings, we first refine a kind of personality-aware word embedding via position encoding and an attention mechanism. Second, to fully fuse knowledge and semantics, we align the personality lexicons with the constructed personality knowledge graph and automatically build a heterogeneous personality word graph for each user. Then, we develop a Heterogeneous Personality Message-passing graph neural Network (HPMN) and perform interactions among the word nodes, emotion and personality heterogeneous nodes in directed edges. Finally, regarding the interpretability of personality traits, we design a graph-level readout function, which delicately selects all heterogeneous nodes for incorporation as knowledge-guided document embeddings to achieve user-generated text personality understanding and interpretation. Therefore, personality detection is transformed into a heterogeneous word graph classification problem. After conducting a verification on 4 public personality datasets, the results show that our model can effectively improve the accuracy of personality detection and pay more attention to critical knowledge.

In summary, our contributions can be summarized as follows.

- To the best of our knowledge, this is the first work that integrates lexical psycholinguistic knowledge and text semantics information into a neural model to achieve interpretable personality detection. Moreover, it provides support for lexical hypotheses in psycholinguistic research from a computational linguistic perspective.
- Our model incorporates the distribution representations of words and the lexical knowledge by learning personality-aware word embeddings. In addition, we construct a heterogeneous personality word graph and develop a message-passing network, which extracts explicit lexicon and knowledge relations via the interactions among heterogeneous graph nodes. All heterogeneous nodes are selectively incorporated as knowledge-guided document embeddings for personality understanding and interpretation through a carefully designed graph readout layer.
- Experiment results on four public datasets demonstrate that our model outperforms the state-of-the-art techniques in terms of personality detection. Our model can help various types of social software mine user information and help psychologists study and analyze personality traits in depth.

The rest of this paper is organized as follows. Section 2 introduces the work related to personality detection. Section 3 provides the problem formulation and Section 4 describes the proposed method. Further, Section 5 presents and analyzes the experimental results obtained on 4 public datasets. Finally, Section 6 outlines the conclusion and future research.

2. Related work

Due to the wide potential application value, personality detection has gradually attracted the attention of computer science researchers [24,25]. Although personality detection in social networks is in its infancy, scholars have achieved fruitful results from multiple research perspectives. Aiming at the challenges mentioned in the previous section, this section focuses on the achievements of scholars in terms of four aspects: (1) psycholinguistic lexicon-based, (2) neural language model-based, (3) user group-based, and (4) interpretability model-based approaches.

2.1. Psycholinguistic lexicon-based methods

Early researchers used language lexicons statistics features such as LIWC [7], Mairesse [26], and MRC [8] features to model personality because they can easily provide insights into the language types that are related to specific personality traits. Tandera et al. extracted features with LIWC from social blogs, and personality traits were predicted by machine learning [9]. Arnoux et al. integrated LIWC and MRC language features with Twitter profile statistics and then trained two machine learning models to predict personality scores [10]. However, these methods are based on the shallow statistical features of vocabulary, without deep psychological knowledge.

2.2. Neural language model-based methods

With the development of deep learning, many natural language processing models have been formed to solve the personality detection task. Xue et al. designed a two-level hierarchical

[1] https://www.mindgarden.com/133-personality-adjective-check-list#horizontalTab6.

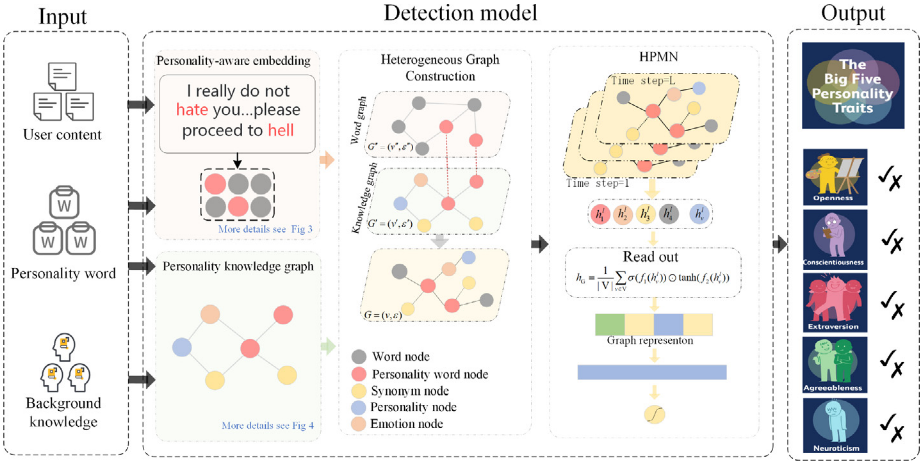


Fig. 2. The overall of personality detection model.

neural network called AttRCNN to learn the deep semantic features of each user's text posts and then concatenated them with LIWC features to predict the Big Five personality scores [11]. Majumder et al. applied 1D convolutions to extract n-grams and combined them with Mairesse features for personality detection [12]. Sun et al. proposed a model called 2CLSTM, which uses the structure of text to detect a user's personality through the combination of bi-directional long short-term memory (Bi-LSTM) and convolutional neural networks (CNN) [13]. Recently, pretraining models have been widely applied to the task of personality detection [10,15], examples include global vectors for word representation (GloVe) [27] and bidirectional encoder representations from transformers (BERT) [28]. The birth of the attention mechanism is of great significance for the understanding of text semantics [29]. Lynn et al. employed a hierarchical attention model to learn the relative weights of users' social media posts to evaluate their personality traits [30]. However, these studies only considered semantics or simply concatenated semantic features with statistical linguistic features without utilizing combinations of semantic information and domain knowledge information.

2.3. User group-based methods

To fit small personality datasets, researchers have explored them from user group perspectives. Network representation learning (NRL) is adopted for the task of group personality detection due to its ability to learn structural features. AdaWalk [31] was the first approach to detect personality via NRL; this method considers the influence of user text-generated networks. Personality2vec is a novel NRL model that makes full use of the semantics, statistics features, and structural information derived from texts to generate a personality vector for each user [32]. Graph neural networks (GNNs) can effectively deal with tasks involving rich relational structures and learn a feature representation for each node in the graph according to the observed structural information [33]. Recently, GNNs have attracted wide attention in works related to different tasks, such as text sentiment analysis [34], rumor detection [35] and knowledge tracing [36]. PersonalityGCN [37] utilize all user information in a heterogeneous graph and firstly employ GNNs [38] to learn users, words, and document embeddings for personality detection. However, the above methods construct a graph from the global structure of the user group, and the test text is essential in the training process. Therefore, these transductive learning methods are not suitable for online personality detection.

2.4. Interpretability model-based methods

Finally, aiming at the interpretability of personality prediction, Poria et al. found that the combination of common knowledge and linguistic features could significantly improve the accuracy of detection [14]. In particular, the authors used SenticNet [39], a popular tool for extracting common knowledge and related emotional polarity from text. In the latest work, Ren et al. proposed a personality detection method combining semantic features and emotional features, increasing the ability to perform personality interpretation and helping to analyze personality traits [15]. Mehta et al. predicted personalities with psycholinguistic and language model features and quantified the impacts of various psycholinguistic statistical features [16]. However, these studies did not effectively integrate psycholinguistic domain knowledge into an automatic personality detection model.

3. Problem formulation

Personality detection can be formulated as a user-level multilabel classification problem. Mathematically, given a user-generated document $\mathrm{D} = \{\mathrm{s}_1, \mathrm{s}_2, \ldots, \mathrm{s}_n\}$, where $\mathrm{s}_i = \{w_i^1, w_i^2, \ldots, w_i^m\}$ is the ith sentence with m words. Our goal is to detect T personality traits $\mathrm{Y} = \{\mathrm{y}^t\}_{t=1}^{T}$ for this user based on document D, where $\mathrm{y}^t \in \{0, 1\}$ is a binary variable.

4. Proposed method

In this section, we present our GNN-based personality detection model guided by lexical psycholinguistic knowledge. Our model takes full advantage of personality lexicons as a bridge to enrich the representations of personality documents with the incorporation of heterogeneous external knowledge. As illustrated in Fig. 2, our model contains three main parts.

(1) Personality-aware word embedding: To fully fuse lexical psycholinguistic knowledge and text semantics, we design personality word position encodings and a dual-level attention mechanism to refine more accurate word embeddings.
(2) Heterogeneous personality word graph construction: To extract explicit lexicon and knowledge relations, a heterogeneous personality word graph is automatically constructed for each user while combining the background knowledge in the field of psycholinguistics.
(3) Heterogeneous personality message-passing graph neural network (HPMN): Message-passing is carried out on the constructed heterogeneous personality word graph, and a graph readout function is designed for interpretable personality detection.

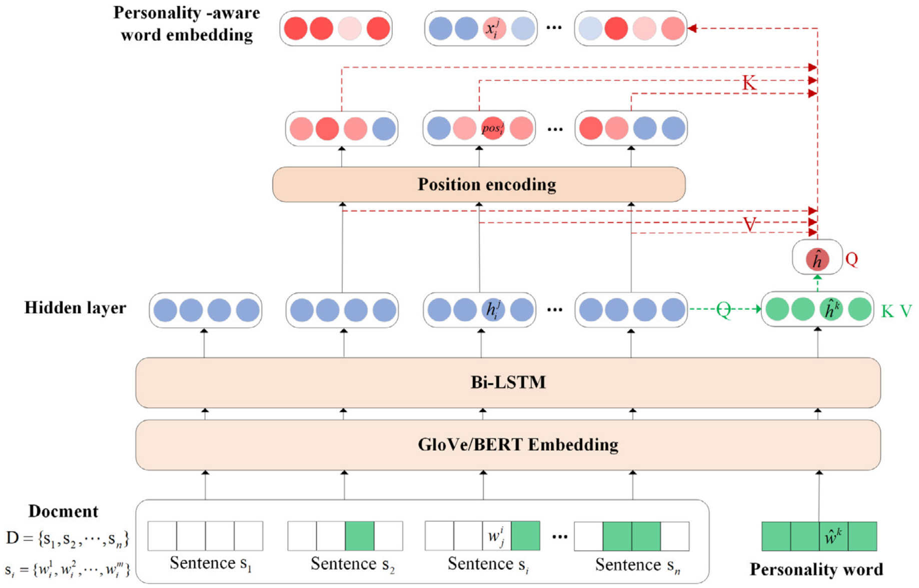


Fig. 3. The personality-aware word embedding process, including the input embedding, position encoding, and dual-attention layers, where the green q, k, and v are the query key and value of the sentence-to-personality word attention, and the red q, k, and v correspond to the personality word-to-sentence attention.

4.1. Personality-aware word embedding

From a psycholinguistic perspective, psychologists believe that language is the most common and reliable way for people to express their inner thoughts and feelings [40]. The personality vocabulary hypothesis provides strong lexical evidence and support for personality detection [20,21]. By combining expert psycholinguistic knowledge, we compile a personality dictionary containing 2043 personality words, including LIWC words and some adjectives. Each word in the dictionary is associated with some personality trait category; in other words, each word contains weakly supervised information. Based on this personality dictionary, the personality-aware word embedding approach is designed in this subsection. The specific steps are shown in Fig. 3, including those related to the input embedding layer, position encoding layer, and dual-attention layer.

4.1.1. Input embedding

In the input embedding layer, the initial embedding of each word in the corpus is derived by pretrained GloVe [27] and BERT [28]. To capture the semantic information of each document, each word w_i^j in sentence $s_i = \{w_i^1, w_i^2, \dots, w_i^m\}$ is fed through a Bi-LSTM to produce a hidden state:

$$h_i^j = \mathrm{Bi-LSTM}(w_i^j), i \in \{1, 2, \dots, \mathrm{n}\}, \mathrm{j} \in \{1, 2, \dots, \mathrm{m}\}, \tag{1}$$

where n and m indicate the numbers of sentences in the document and words in the sentence, respectively. Combined with the personality dictionary $\widehat{W} = \{\hat{w}^1, \hat{w}^2, \dots, \hat{w}^K\}$, for each personality word $\hat{w}^k$ in the document, we also employ Bi-LSTM to capture the sequential information contained in the personality words of the full document:

$$\hat{h}^k = \mathrm{Bi-LSTM}(\hat{w}^k) \tag{2}$$

4.1.2. Position encoding

Implicit semantic relationships are present between personality words and neighbor words in text. Inspired by the position encoding approach used in the literature [41], we define a position index to model the position information of the personality words in the corresponding sentence. Specifically, suppose we are given a sentence $s_i = \{w_i^1, w_i^2, \dots, w_i^m\}$ with m words, which contain n personality words; $0 < n \leqslant m$. The relative distance between the pth word and the qth personality word is defined as follows:

$$dis_{w_i^q}^{w_i^p} = \begin{cases} 1 - \dfrac{|index(p) - index(q)|}{m}, if(n \neq m) \\ 1, if(n = m) \end{cases}, \tag{3}$$

where $index(p)$ and $index(q)$ indicate the position indices of the non-personality word and the personality word in the sentence, respectively. For example, in the first sentence "*I really don't hate you*" in Fig. 1, "*hate*" is a personality word, and its position code is $pos = [0.4, 0.6, 0.8, 1, 0.8]$. By looking up this position code, the representation of sentence s_i with personality position information is expressed as:

$$pos_i^j = \frac{1}{n}\sum_{n=0}^{n} dis_{w_i^q}^{w_i^p} \times h_i^j, \tag{4}$$

$$s_i' = [pos_i^1, pos_i^2, \dots, pos_i^m], \tag{5}$$

where h_i^j is the hidden state of the jth word and pos_i^j is the position code of the jth word obtained after considering the n personality words in sentence s_i. Note that no personality words are contained in the sentence and no positional encoding is required.

4.1.3. Dual-attention

To capture the interactive information between sentence-level semantics and document-level personality words, a dual-attention mechanism is designed to refine the representations of words. Intuitively, personality words possess not only global document semantics but also local sentence semantics. Thus, we first devise a sentence-to-personality word attention to merge the representations of sentence s_i into a personality word representation. The definitions of sentence-to-personality word attention are as follows:

$$e_{sp} = Avg(h_i^j)^{\mathrm{T}} \cdot \mathrm{W}_{sp} \cdot \hat{h}^k, \tag{6}$$

$$\alpha_k = \frac{\exp(e_{sp})}{\sum_{k=1}^{K} \exp(e_{sp})}, \tag{7}$$

$$\hat{h} = \sum_{k=1}^{K} \alpha_k \cdot \hat{h}^k, \tag{8}$$

where W_{sp} and $Avg(h_i^j)$ are the trainable weight matrix and average pooling operation of the document, respectively. α_k is the attention weight for the corresponding personality word sequence, and $\hat{h}$ denotes the new representations of personality words under the attention of the sentence semantics.

Based on the above personality words, we then devise a personality-to-sentence attention to learn the personality-aware word embedding x_i^j in each sentence:

$$e_{ps} = \hat{h}^{\mathrm{T}} \cdot \mathrm{W}_{ps} \cdot pos_i^j, \tag{9}$$

$$\beta_j = \frac{\exp(e_{ps})}{\sum_{j=1}^{m} \exp(e_{ps})}, \tag{10}$$

$$x_i^j = \beta_j \cdot h_i^j, i \in \{1, 2, \ldots, \mathrm{n}\}, \mathrm{j} \in \{1, 2, \ldots, \mathrm{m}\}, \tag{11}$$

where the new personality word representation $\hat{h}$ denotes a query, each hidden state after a position encoding pos_i^j denotes a key, and each hidden state h_i^j of a sentence s_i is used as a value. β_j is the attention weight for the corresponding personality-aware word embedding in each sentence. x_i^j denotes the personality-aware representation of the jth word in the ith sentence in a document.

4.2. Heterogeneous personality word graph construction

To better fuse knowledge and semantics, we model a user-generated document as a heterogeneous graph over word nodes and knowledge nodes. Specifically, a personality knowledge graph G' is constructed based on word-level psycholinguistic knowledge. Then, a word graph is constructed for each user-generated document G'' based on word co-occurrence. Finally, a heterogeneous personality graph G is generated for each user via personality word node alignment in the word graph and knowledge graph.

4.2.1. Personality knowledge graph

First, we summarize the results of psycholinguistic research and express them as a personality knowledge graph $G' = (\upsilon', \varepsilon')$ with four types of nodes and three relationships. As shown in Fig. 4, the personality knowledge graph describes the symbiotic relationships between personality words, synonyms, personality entities, and emotional entities. Specifically, the personality words include 2043 words, which are manually organized based on the psychological research in the appendix. Personality synonyms are the synonyms of the top-k selected personality words. SenticNet [39] is an emotional knowledge base that contains rich emotional attributes, which provide conceptual representations of the emotions of words. Based on the research results regarding the relationships between emotions and personality traits in the literature [23], we manually construct symbiotic relationships between the emotion attributes and personality traits in the graph. By combining the relationships between the four types of nodes, we construct a word-level personality knowledge graph that reflects the theoretical knowledge of personality linguistics.

4.2.2. Word graph

Then, we construct a word graph for each user-generated document $G'' = (\upsilon'', \varepsilon'')$ through the unique representations of word vertices and word co-occurrences. Specifically, a fixed-size sliding window (the default length is 3) is used to determine the co-occurrence information of words. Point mutual information (PMI) [38] is used to calculate the connection weight between two words. A positive PMI value indicates high semantic similarity between the tested two words, while a negative PMI value indicates weak or even irrelevant similarity. For example, the adjacency matrix of the document "*I really don't hate you...please proceed to hell*" is shown in Fig. 5(a). The initial representation of the word node is a personality-aware embedding x_i^j.

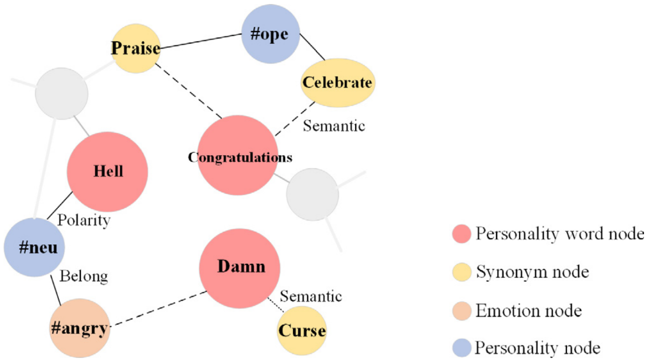


Fig. 4. Personality knowledge graph.

4.2.3. Heterogeneous personality word graph

By align the personality word nodes in the knowledge graph G' and word graph G'', we construct a heterogeneous personality word graph $G = (\upsilon, \varepsilon)$ for each user. υ contains the word node in a user-generated document and the entity node in the sampled knowledge subgraph. ε denotes the relationships between multiple types of nodes. For instance, the heterogeneous graph adjacency matrix of the document "*I really don't hate you...please proceed to hell*" is shown in Fig. 5(b). The initial representation of the emotional entity nodes and personality entity nodes is the term frequency-inverse document frequency (TF-IDF) vector of its Wikipedia corpus.

4.3. HPMN

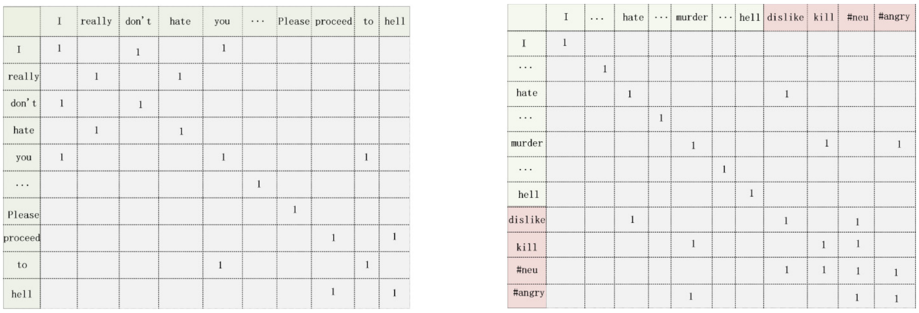
Upon constructing the heterogeneous personality word graph of each user, we propose an HPMN to fully exploit the hidden interactions between words and psycholinguistics knowledge, and the personality detection task is then transformed into a heterogeneous word graph classification problem. As shown in Fig. 2, the HPMN is composed of two parts: heterogeneous graph interaction and graph-level readout modules. Message passing is carried out in the constructed heterogeneous personality word graph to learn the embeddings of the nodes. The graph-level readout module pays attention to the key nodes, which can explain the composition of personality language in terms of word granularity.

4.3.1. Heterogeneous graph interaction

Inspired by TextING [42], our HPMN is a recurrent neural network for processing graph data that iteratively propagates the information of all kinds of nodes through graphs. We employ the gated graph neural networks (GGNN) [43] to learn the embeddings of the nodes on each heterogeneous word graph. A node can aggregate information from its adjacent neighbors through a gating mechanism at each time step. This process is expressed as:

$$a_v^l = A_v[h_1^{l-1}, \ldots, h_{|V|}^{l-1}] + b, \tag{12}$$

where |V| denotes the numbers of all kinds of nodes in the heterogeneous personality word graph G and A_v denotes the



	I	really	don’t	hate	you	···	Please	proceed	to	hell
I	1		1		1					
really		1		1						
don’t	1		1							
hate		1		1						
you	1				1				1	
···						1				
Please							1			
proceed								1		1
to					1				1	
hell								1		1

(a) Word graph adjacency matrix

	I	···	hate	···	murder	···	hell	dislike	kill	#neu	#angry
I	1										
···		1									
hate			1					1			
···				1							
murder					1				1		1
···						1					
hell							1				
dislike			1					1		1	
kill					1				1	1	
#neu								1	1	1	1
#angry					1					1	1

(b) Heterogeneous graph adjacency matrix

Fig. 5. Graph based word-knowledge relation.

subadjacency matrix of node v. As the graph layer operates on the first-order neighbors, high-order feature interaction can be realized by stacking the HPMN layers L times, and one node can reach another node that is L hops away. The interaction formula is:

$$z_v^l = \sigma(W_z a_i^l + U_z h_v^{l-1}), \tag{13}$$

$$r_v^l = \sigma(W_r a_i^l + U_r h_v^{l-1}), \tag{14}$$

$$\tilde{h}_v^l = \tanh(W_h a_v^l + U_h(r_v^l \odot h_v^{l-1})), \tag{15}$$

$$h_v^l = (1 - z_v^l) \odot h_v^{l-1} + z_v^l \odot \tilde{h}_v^l, \tag{16}$$

where σ denotes the sigmoid function, $\odot$ denotes the element-wise multiplication operation, and all W, U and b are trainable weights and biases. z and r are update gates and reset gates used to determine the contribution of neighbor information to the current node embedding, respectively.

4.3.2. *Readout function*

After going through the L-layer HPMN, all kinds of nodes are sufficiently updated. We design a new readout function to aggregate the nodes into graph-level representations for documents. The specific readout function is:

$$h_G = \frac{1}{|V|} \sum_{v \in V} \sigma(f_1(h_v^l)) \odot \tanh(f_2(h_v^l)), \tag{17}$$

where f_1 is a soft attention mechanism that decides which nodes are relevant to the current graph-level task. f_2 is a multilayer perceptron that is used as a nonlinear feature transformation.

Finally, we employ T sigmoid-normalized linear transformations on the graph embeddings h_G to detect T personality traits. For the tth personality trait, the corresponding probabilities can be calculated as:

$$p(y^t) = sigmoid(\mathrm{W}_t h_G + b_t), \tag{18}$$

where W_t is a trainable weight matrix and b_t is a bias term. The loss function is defined by the binary cross-entropy of the predicted and true label distributions during training:

$$\mathcal{L} = -y_t \log p(y^t) + (1 - y_t) \log(1 - p(y^t)), \tag{19}$$

where $p(y^t)$ is the predicted probability for the tth personality trait of a user and y_t is the ground truth of this personality trait.

4.4. *Learning algorithm*

In summary, this paper mainly designs two algorithms: the personality-aware word embedding algorithm 1 and the HPMN algorithm 2. Based on a personality dictionary, the personality-aware word embedding algorithm is designed to refine more accurate word embeddings. Then, by aligning the personality word nodes in the knowledge graph and word graph, a heterogeneous personality word graph is constructed for each user. In the HPMN algorithm, message passing is carried out on the constructed heterogeneous personality word graph, and a graph readout function is designed to achieve interpretable personality detection. The experiment mainly includes model training and model testing.

Algorithm 1 Personality-aware word embedding algorithm

```
Input:
  A user-generated document $D = \{s_1, s_2, \cdots, s_n\}, s_i = \{w_i^1, w_i^2, \cdots, w_i^m\}$;
  Personality dictionary $\widehat{W} = \{\hat{w}^1, \hat{w}^2, \cdots, \hat{w}^K\}$;
Output:
  Personality-aware word embeddings $x_i^j, i \in \{1, 2, \cdots, n\}, j \in \{1, 2, \cdots, m\}$;
  //Input embeddings
  Initialize word embeddings
  Obtain the contextual embedding of each word $h_i^j$ and personality word $\hat{h}^k$ according to Eqs (1)–(2);
  //Position encodings
  for each word $w_i^j$ in sentence $s_i$ do
    Calculate the relative personality word distance $dis_{w_i^q}^{w_i^p}$ according to Eq. (3);
    Obtain the representation of sentence $s_i$ with personality position information $s_i{}'$ according to Eqs. (4)–(5);
  end for
  //Dual-attention mechanism
  for each personality word $\hat{w}^k$ in document $D$ do
    Calculate sentence-to-personality attention values for words to obtain new representations of the personality words $\hat{h}$ according to Eqs. (6)-(8);
  end for
  for each sentence $s_i$ in document $D$ do
    Calculate personality word-to-sentence attention to learn the personality-aware word embeddings $x_i^j$ according to Eqs. (9)–(11);
  end for
  return Personality-aware word embedding $x_i^j$;
```

Algorithm 2 HPMN personality detection algorithm

```
Input:
    Personality knowledge graph $G' = (\upsilon', \varepsilon')$;
    Word graph of each user-generated document $G'' = (\upsilon'', \varepsilon'')$;
Output:
    Personality trait detection results $\mathrm{Y} = \{\mathrm{y}^t\}_{t=1}^T$;
    // Initialization and construction
    Initialize word embedding $x_i^j$ from algorithm 1;
    Initialize all trainable weights and biases $W$, $U$, and $b$;
    Construct a heterogeneous personality word graph for each user $G = (\upsilon, \varepsilon)$;
    // Model training
    for training data do
        Stack $L$ heterogeneous message-passing layers to capture the interactions on each heterogeneous word graph according to Eqs. (12)–(16);
        Obtain the representation of each heterogeneous word graph $h_G$ through the readout function according to Eq. (17);
        Calculate the probabilities of $T$ personality traits via sigmoid-normalized linear transformations on the graph embeddings $h_G$ according to Eq. (18);
    end for
    repeat
        update all trainable weights and biases $W$, $U$, and $b$;
    until converge
    // Model testing
    for test data do
        Detect $T$ traits $Y = argmaxp(y^t|G), t \in \{1, 2, \cdots, T\}$ for each user;
    end for
```

5. Experiments and analysis

5.1. Experimental settings

In this section, we introduce the datasets used in the experiment and present the baseline methods. After introducing the parameter set, we present the evaluation index used to evaluate the performance of the models.

5.1.1. Datasets

To date, most research on personality analysis has used the Big Five model of personality, where each trait is represented by a continuous score. We choose 4 public benchmark datasets for the experiment, and the Big Five regression scores in the MyPersonality YouTube and PAN datasets are converted into class labels for different traits. Without processing, the original labels are used on the Essays dataset. The details of the datasets are as follows.

MyPersonality[2]: The MyPersonality project is a Facebook App that allowed its users to participate in psychological research by filling out a personality questionnaire. This dataset contains 9917 status updates of 250 users and their Big Five scores.

Essays[44]: This dataset consists of 2468 anonymous essays tagged with the authors personality traits. Stream-of-consciousness essays were written by volunteers in a controlled environment, and the authors of the essays were asked to label their Big Five personality traits. We remove one essay that contains only the text "Err:508" from the dataset, and we experiment with the remaining 2467 essays.

YouTube[3]: This dataset consists of a collection of speech transcriptions provided by 404 users with their Big Five personality scores. The labels of this dataset were collected from the crowd-sourced annotation task. Annotators watched each video blog and then rated the corresponding Big Five personality scores with a questionnaire.

PAN[4]: This dataset was collected from the PAN2015 data science competition. It consists of a four language dataset, and we choose the English dataset, which contains 294 users Twitter tweets and their Big Five scores.

5.1.2. Baseline methods

To verify the performance of the proposed method, we compare our approach with some baseline methods on different datasets.

CNN [12]: This method applies 1D convolutions to extract n-grams in combination with Mairesse features for personality detection.

AdWalk [13]: This is the first method to introduce NRL into personality analysis. It calculates the text similarities between users and constructs the user graph, and then designs a walking strategy based on node2vec to learn the user embeddings and analyze their personalities.

Personality2vec [32]: This method makes full use of the semantics, personality, and structural information in user texts and designs an NRL model. This model utilizes a new biased walk algorithm and an improved skip-gram algorithm to conduct training on the graph and eventually generates a personality vector for each user node.

PersonalityGCN [37]: This method constructs a heterogeneous graph of user document and applies a text graph convolutional network for personality detection.

Hierarchical Model [30]: This method uses hierarchical attention to encode user messages into a representation that can be used to predict the personality of the user.

BERT+emotion [15]: This is an automatic personality detection model based on a neural network that combines emotional features and semantic features. To the best of our knowledge, this is the state-of-the-art technique for use with the Essays and MBTI datasets developed to date.

Next, we introduce the experimental settings of the baseline methods described above. We use the default hyperparameters for the other methods as employed in the related papers or source codes. However, we make some modifications to the different baselines. For the AdWalk and Personality2vec models, we replace the support vector regression (SVR) module with fully connected and sigmoid layers to change the regression task into a multilabel classification task and add an experiment on the Essays dataset. For the PersonalityGCN model, in addition to the author's experiments on the MyPersonality and Essays datasets, we also complete experiments on the remaining datasets with this baseline method. For the hierarchical model, we use a pretrained BERT model to initialize the 200-dimensional word embeddings. Different from the original model, we feed two hidden layers and a sigmoid layer for multilabel classification.

5.1.3. Parameter set and evaluation metrics

For all datasets, we randomly divide all data at a ratio of 8:1:1 for the actual training set, validation set and test set, respectively. The hyperparameters are adjusted based on the performance achieved on the validation set. Empirically, we set the batch size to 32, the learning rate to 0.001 (with the Adam optimizer), and

[2] http://mypersonality.org.

[3] https://www.idiap.ch/en/dataset/youtube-personality/index_html.

[4] https://pan.webis.de/clef15/pan15-web/author-profiling.html.

Table 1
Comparison with baselines and ablated models in 4 datasets.

Dataset	Model		Traits					
			EXT	NEU	AGR	CON	OPN	Average
MyPersonality	Baselines	CNN	–	–	–	–	–	–
		AdWalk	0.624	0.616	0.634	0.619	0.645	0.627
		Personality2vec	0.673	0.642	0.658	0.632	0.670	0.655
		Personality GCN	0.800	0.790	0.680	0.760	0.800	0.766
		Hierarchical model	0.780	0.761	0.678	0.793	0.745	0.751
		BERT+emotion	–	–	–	–	–	–
	Ablated models	PMN (GloVe)	0.804	0.774	0.694	0.788	0.759	0.763
		PMN (BERT)	0.820	0.794	0.711	0.798	0.789	0.782
	Full models	HPMN (GloVe)	0.811	0.813	0.707	0.816	0.802	0.789
		HPMN (BERT)	**0.825**	**0.827**	**0.727**	**0.826**	**0.823**	**0.805**
Essays	Baselines	CNN	0.581	0.594	0.567	0.573	0.627	0.588
		AdWalk	0.616	0.624	0.584	0.599	0.637	0.612
		Personality2vec	0.636	0.612	0.594	0.619	0.663	0.624
		Personality GCN	0.600	0.630	0.577	0.591	0.648	0.609
		Hierarchical model	0.713	0.764	0.634	0.743	0.738	0.718
		BERT+emotion	0.799	0.801	0.803	**0.802**	0.804	0.801
	Ablated models	PMN (GloVe)	0.734	0.773	0.752	0.730	0.769	0.751
		PMN (BERT)	0.750	0.794	0.741	0.763	0.790	0.767
	Full models	HPMN (GloVe)	0.737	0.778	0.780	0.786	0.803	0.776
		HPMN (BERT)	**0.811**	**0.817**	**0.807**	0.796	**0.818**	**0.809**
YouTube	Baselines	CNN	–	–	–	–	–	–
		AdWalk	0.616	0.656	0.624	0.699	0.665	0.652
		Personality2vec	0.633	0.673	0.614	0.681	0.660	0.652
		Personality GCN	0.680	0.708	0.609	0.693	0.687	0.675
		Hierarchical model	0.662	0.695	0.628	0.71	0.652	0.669
		BERT+emotion	–	–	–	–	–	–
	Ablated models	PMN (GloVe)	0.693	0.750	0.651	0.708	0.710	0.702
		PMN (BERT)	0.721	0.780	0.676	0.710	0.729	0.723
	Full models	HPMN (GloVe)	0.735	0.813	0.697	0.726	0.753	0.744
		HPMN (BERT)	**0.755**	**0.823**	**0.721**	**0.763**	**0.773**	**0.767**
PAN	Baselines	CNN	–	–	–	–	–	–
		AdWalk	0.621	0.670	0.625	0.631	0.629	0.635
		Personality2vec	0.675	0.636	0.641	0.619	0.655	0.645
		Personality GCN	0.586	0.561	0.564	0.566	0.587	0.572
		Hierarchical model	0.613	0.605	0.598	0.623	0.625	0.612
		BERT+emotion	–	–	–	–	–	–
	Ablated models	PMN (GloVe)	0.574	0.601	0.584	0.588	0.589	0.587
		PMN (BERT)	0.609	0.634	0.603	0.593	0.609	0.609
	Full models	HPMN (GloVe)	0.663	0.683	0.647	0.631	0.628	0.650
		HPMN (BERT)	**0.688**	**0.713**	**0.663**	**0.646**	**0.668**	**0.675**

the dropout rate to 0.5. The pretrained language models GloVe[5] and BERT are employed to initialize the word embeddings. The network structure of the uncased BERT-based model[6] contains 12 layers, 768 hidden layers, and 12 heads. The dimensions of the word vectors of GloVe and BERT are 300 and 768, respectively. The out-of-vocabulary words are randomly sampled from a uniform distribution [-0.01, 0.01]. The number of Bi-LSTM hidden units is set to 300. For a fair comparison, we use the same embedding in other ablated models.

In this paper, we introduce the accuracy and F1-measure to evaluate the detection results. The F1-measure is the harmonic average of precision and recall. The closer this measure is to 1, the better the detection effect of the corresponding algorithm. Specifically, this paper reports the highest accuracy and F1-score of each trait achieved for each dataset by each method.

5.2. Performance analysis

In this section, the performance of the proposed method is evaluated in 4 sets of experiments. First, the effectiveness of our method is verified by comparison with other baseline methods. Subsequently, the time efficiency analysis is performed. Then, the personality-aware embedding method is validated. Furthermore, the parameter sensitivity analysis are presented. Finally, the interpretability of our approach regarding how words impact personality detection is presented.

5.2.1. Detection performance analysis

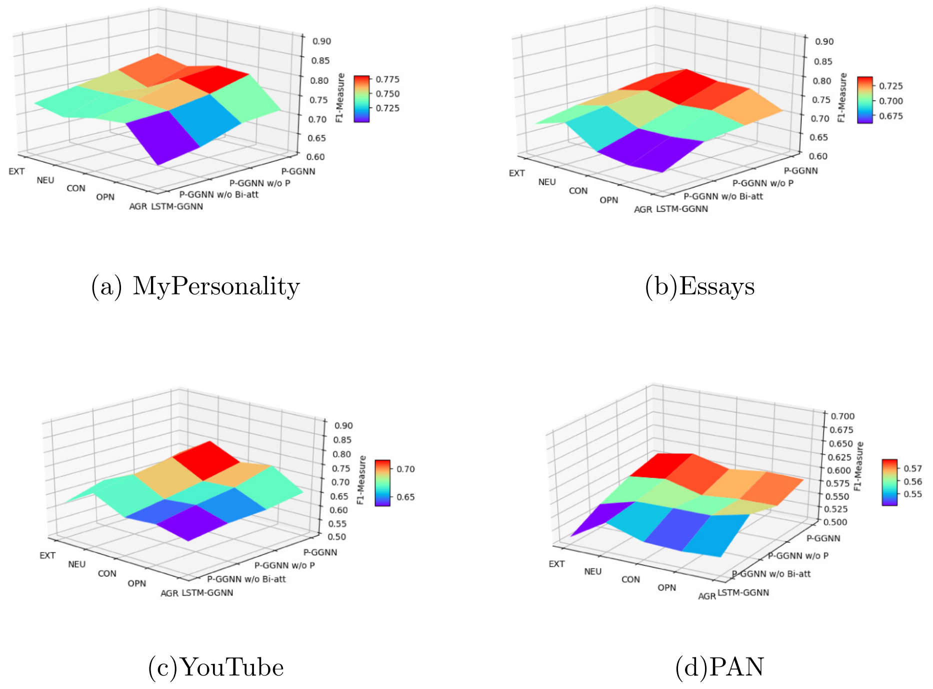
As shown in Table 1, we report the accuracy of the classic baselines, our proposed HPMN model, and ablations with GloVe embeddings or BERT embeddings on 4 public datasets. Based on the Table 1, we obtain the following observations.

A comprehensive comparison among all baselines on 4 public datasets shows that the average accuracy rate of our approach is increased by 7.26%. Compared with the state-of-the-art techniques (BERT + emotion), HPMN obtains an average accuracy improvement of 1% on the Essays dataset. This shows that our model has good detection ability. This is because we merge the vertical field of effective external knowledge to supplement the input text information and carry more information, as discussed in their future work.

On the PAN dataset, the traditional feature-based and neural network methods such as PersonalityGCN and the hierarchical models have poor overall performance. It is difficult for them to

[5] http://nlp.stanford.edu/data/glove.6B.zip.

[6] https://storage.googleapis.com/bert_models/2018_10_18/uncased_L-12_H-768_A-12.zip.



(a) MyPersonality (b)Essays

(c)YouTube (d)PAN

Fig. 6. F1-score achieved by the ablation models.

catch enough information during the learning process since the text length of the PAN dataset is shorter than the other datasets. The NRL methods (AdWalk, Personality2vec) are competitive on the PAN dataset. This is because they detect personality from the perspective of the user group, which is not applied to online personality detection. Our HPMN has good detection performance due to the extra information carried by the text, and it can detect personality online via inductive learning.

On the 4 datasets, we can observe that the feature of EXT and NEU traits can be effectively extracted from the text. In other words, the feature information of the AGR personality trait in the text information is not obvious. We hypothesize that this is because people with NEU and EXT traits often use modal particles and adjectives that are more recognizable, while people with AGR traits have a more peaceful tone.

Compared with the ablation model PMN on the 4 datasets, the full HPMN model performs better. These observations indicate that our HPMN can enrich the representations of personalities derived from user content by iterating the observed semantic information and knowledge. This demonstrates the feasibility of applying lexical psycholinguistic knowledge to facilitate personality detection from text. The BERT-based model performs better than the GloVe-based model in most cases, which shows that the pretrained BERT model has a positive effect on the personality classification task.

5.2.2. Time efficiency analysis

In practice, a user personality detection system should have both high accuracy and low computational complexity to meet actual requirements. Therefore, it is essential to analyze the time efficiency of a personality detection model as well. Table 2 shows the time consumption of the proposed model compared to the baselines. The values given in Table 2 are the average time costs achieved for detecting users in testset. All experiments run on a server consisting of a 3.4 GHz Intel Xeon E5-2620 v4 CPU with 62 GB of RAM and three Nvidia GeForce RTX310 2080ti GPUs. For a fair comparison, the time consumption of data pre-processing is not considered.

Table 2
Time consumption of different methods on 4 datasets (milliseconds/user).

Model	MyPrsonality	Essays	YouTube	PAN
CNN	–	2.22	–	–
AdWalk	13 155.4	40 579.2	21 473.6	15 682.0
Personality2vec	18 711.8	43 367.3	17 265.6	16 373.7
Personality GCN	66 384.4	80 446.6	76 501.4	52 158.3
Hierarchical model	3.68	3.52	3.67	3.75
BERT+emotion	–	4.53	–	–
HPMN	5.07	5.69	5.79	5.89

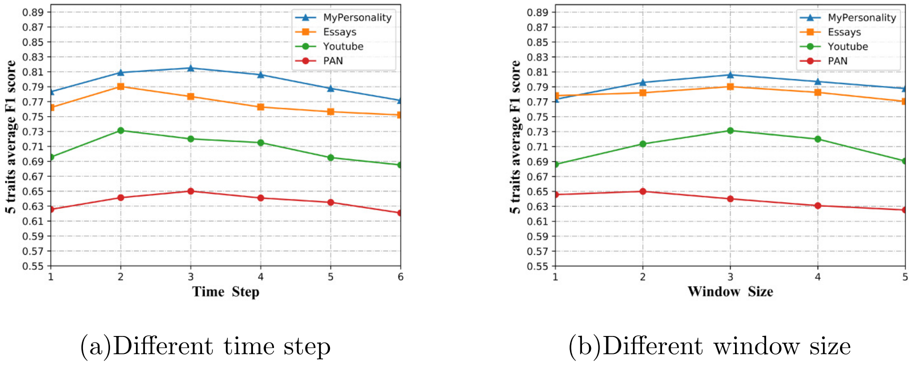
Table 3
The recurrence ratio between the top-10 identified high-attention words and personality words in the test sets of MyPersonality and Essays.

Model	MyPrsonality	Essays
HEU	75%	73%
OPE	68%	71%
CON	58%	63%
EXT	70%	72%
AGR	56%	58%

As we can observe in Table 2, AdWalk, Personality2vec and Personality GCN require high time consumption levels to implement personality detection. This is because they are essentially solving a semi-supervised graph node classification task, so these models need to be retrained when testing on new data. Our proposed HPMN achieves the fourth-lowest running time cost with a small difference from the optimal CNN model. This is due to the slightly higher number of parameters in our multi-layer message aggregation than in the multi-layer perceptron (MLP). However, considering the more accurate detection results achieved, these small time consumption gaps are negligible in real systems. Therefore, the complexity of our HPMN is acceptable.

5.2.3. Analysis of personality-aware embedding

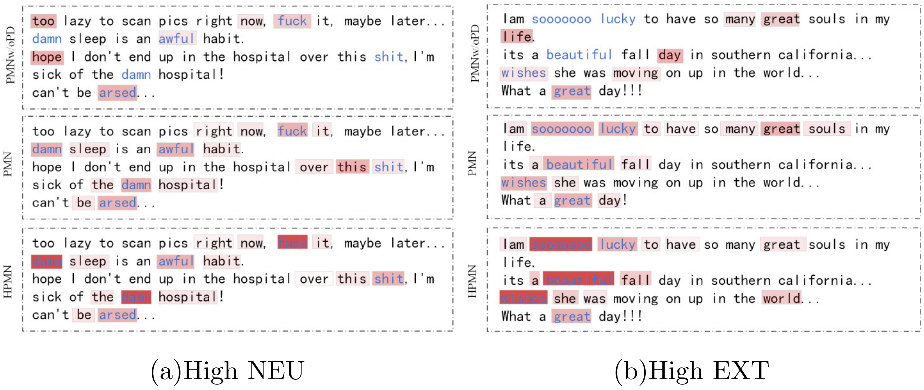
To verify the effectiveness of the proposed personality-aware word embedding method, we ablate the positional encoding module (PMN w/o P), dual-attention mechanism (PMN w/o D),



(a)Different time step

(b)Different window size

Fig. 7. F1-score achieved for 5 personality traits with the HPMN (BERT) on MyPersonality dataset.



too lazy to scan pics right now, fuck it, maybe later...
damn sleep is an awful habit.
hope I don't end up in the hospital over this shit, I'm
sick of the damn hospital!
can't be arsed...

too lazy to scan pics right now, fuck it, maybe later...
damn sleep is an awful habit.
hope I don't end up in the hospital over this shit, I'm
sick of the damn hospital!
can't be arsed...

too lazy to scan pics right now, fuck it, maybe later...
damn sleep is an awful habit.
hope I don't end up in the hospital over this shit, I'm
sick of the damn hospital!
can't be arsed...

Iam soooooooo lucky to have so many great souls in my
life.
its a beautiful fall day in southern california...
wishes she was moving on up in the world...
What a great day!!!

Iam soooooooo lucky to have so many great souls in my
life.
its a beautiful fall day in southern california...
wishes she was moving on up in the world...
What a great day!

Iam soooooooo lucky to have so many great souls in my
life.
its a beautiful fall day in southern california...
wishes she was moving on up in the world...
What a great day!!!

(a)High NEU

(b)High EXT

Fig. 8. Illustration of the attention weights obtained by PMN w/o PD, PMN, and HPMN. (a) and (b) are examples of high neuroticism and high extraversion from the MyPersonality dataset. (a) Personality words: fuck, damn, awful, shit, arsed; (b) Personality words: sooooooo, lucky, beautiful, wish, great.

Table 4
Top correlations between the Big Five and individual words.

Trait	No. of words	Top 20 words
Neuroticism	24	Awful (0.26), though (0.24), lazy (0.24), worse (0.21), depressing (0.21), irony (0.21), road (0.2), terrible (0.2), Southern (0.2), stressful (0.19), horrible (0.19), sort (0.19), visited (0.19), annoying (0.19), ashamed (0.19), ground (0.19), ban (0.18), oldest (0.18), invited (0.18), completed (0.18)
Extraversion	20	Bar (0.23), other (0.22), drinks (0.21), restaurant (0.21), dancing (0.2), restaurants (0.2), cats (0.2), grandfather (0.2), Miami (0.2), countless (0.2), drinking (0.19), shots (0.19), computer (0.19), girls (0.19), glorious (0.19), minor (0.19), pool (0.18), crowd (0.18), sang (0.18), grilled (0.18)
Openness	393	Folk (0.32), humans (0.31), of (0.29), poet (0.29), art (0.29), by (0.28), universe (0.28), poetry (0.28), narrative (0.28), culture (0.28), giveaway (0.28), century (0.28), sexual (0.27), films (0.27), novel (0.27), decades (0.27), ink (0.27), passage (0.27), literature (0.27), blues (0.26)
Agreeableness	110	Wonderful (0.28), together (0.26), visiting (0.26), morning (0.26), spring (0.25), porn (0.25), walked (0.23), beautiful (0.23), staying (0.23), felt (0.23), cost (0.23), share (0.23), gray (0.22), joy (0.22), afternoon (0.22), day (0.22), moments (0.22), hug (0.22), glad (0.22), fuck (0.22)
Conscientiousness	13	Completed (0.25), adventure (0.22), stupid (0.22), boring (0.22), adventures (0.2), desperate (0.2), enjoying (0.2), saying (0.2), Hawaii (0.19), utter (0.19), it is (0.19), extreme (0.19), deck (0.18)

and both of the above (PMN w/o PD). According to the 3D trend charts 6 obtained for the 4 datasets, we can infer that the position information encoding module and the dual-attention mechanism are useful components for achieving performance improvements.

5.2.4. Parameter sensitivity

To evaluate how the parameters in the HPMN affect personality detection and to provide references for parameter selection in practice, we conduct a parameter sensitivity analysis. Fig. 7(a) shows the performance of the HPMN (BERT) with a varying number of time steps on MyPersonality, Essays, YouTube, and PAN. The results show that the effect is basically optimal when the number of time steps is 3 or 2. This is because more neighbor nodes can be used to learn a more accurate node representation. However, as the number of layers increases, this situation is reversed. This is because when multiple layers of the HPMN are stacked after a node receives a message from each node in the whole graph, the nodes in the graph become too smooth. Fig. 7(b) shows the detection performance achieved on the four datasets under different word co-occurrence window sizes. This figure shows a trend similar to that observed when the time

Table 5
Top category and word-level correlations for the lower-order facets.

Neuroticism trait	Top 20 words
Anxiety	Awful (0.29), sick (0.26), road (0.26), ground (0.25), terribly (0.25), cranky (0.25), stress (0.24), feeling (0.24), southern (0.24), stressful (0.24), myself (0.23), though (0.23), feel (0.23), sweater (0.23), county (0.23), scenario (0.23), ashamed (0.22), feels (0.22), oldest (0.22), spoiled (0.22)
Anger	Sick (0.24), later (0.23), yay (0.22), road (0.22), possibly (0.22), completely (0.21), 30 (0.21), though (0.21), poem (0.21), wild (0.21), desperately (0.2), pregnancy (0.2), should not (0.2)
Depression	Lazy (0.24), refuse (0.23), irony (0.22), pretend (0.22), visited (0.22), horrible (0.22), harsh (0.22), combined (0.21), stupid (0.21), uncomfortable (0.21), though (0.21), fuck (0.2), drugs (0.2), guardian (0.2)
Self-consciousness	Sizes (0.27), smoke (0.26), city (0.25), Irish (0.24), messy (0.24), football (0.24), wife (0.24), silly (0.24), street (0.23), easier (0.23), opinions (0.23), lazy (0.23), shorter (0.23), expecting (0.23), mountain (0.22), fit (0.22), al (0.22), instead (0.22), realistic (0.22), fire (0.22)
Immoderation	Apart (0.21), drops (0.21), already (0.21)
Vulnerability	Lazy (0.26), awful (0.22), bull (0.22), Southern (0.22), al (0.22), uncomfortable (0.22), lately (0.22), myself (0.21), though (0.21), sunset (0.21), drop (0.21), combined (0.21), feeling (0.2)
Extraversion trait	Top 20 words
Friendliness	Sang (0.27), hotel (0.26), lazy (0.26), kissed (0.26), shots (0.26), golden (0.24), dad (0.24), girls (0.24), restaurant (0.24), eve (0.23), best (0.23), proud (0.23), miss (0.23), accept (0.23), soccer (0.23), met (0.22), not (0.22), brothers (0.22), interest (0.22), cheers (0.22)
Gregariousness	Friends (0.32), girls (0.31), tickets (0.29), Friday (0.28), concert (0.27), enough (0.27), beings (0.27), rather (0.27), drinks (0.27), Ryan (0.27), useful (0.26), ticket (0.26), aka (0.26), birds (0.25), pages (0.25), met (0.25), gentle (0.25), patterns (0.25), haha (0.25), concept (0.25)
Assertiveness	Aka (0.27), countless (0.25), restaurants (0.23), bar (0.21), ticket (0.2), request (0.2)
Activity level	Contrary (0.25), run (0.24), dolls (0.22), for. (0.22), pack (0.22), hours (0.21), 8 (0.21), fiction (0.21), child (0.2)
Excitementseeking	Cats (0.28), football (0.27), sizes (0.27), books (0.27), sewing (0.26), box (0.26), winter (0.25), leaf (0.25), knitting (0.25), blankets (0.25), delightful (0.24), book (0.24), piles (0.24), I am (0.24), haha (0.24), shelf (0.24), asking (0.24), terrific (0.24), gentle (0.24), cat (0.24)
Cheerfulness	Checking (0.27), excitement (0.26), love (0.25), kidding (0.25), hot (0.25), friends (0.25), spend (0.24), shots (0.24), glory (0.23), miss (0.23), sing (0.23), girls (0.23), perfect (0.23), denied (0.23), sweet (0.23), song (0.23), every (0.22), temporary (0.22), dance (0.22), golden (0.22)
Openness trait	Top 20 words
Imagination	Novel (0.29), fame (0.28), urge (0.28), decades (0.27), urban (0.27), 8th (0.26), glance (0.26), length (0.26), poetry (0.26), literature (0.26), audience (0.26), 8 (0.25), anniversary (0.25),6 (0.25), loves (0.25), narrative (0.25), lines (0.24), bears (0.24), thank (0.24), humans (0.24)
Artistic interests	Beauty (0.26), moon (0.26), blues (0.26), sky (0.26), plants (0.26), dance (0.26), beautiful (0.25), trees (0.25), planted (0.25), flowers (0.25), sang (0.25), blue (0.25), sings (0.25), danced (0.25), music (0.24), afterwards (0.24), tree (0.24), painted (0.24), hills (0.24), outdoor (0.23)
Emotionality	Feel (0.29), breathe (0.29), feeling (0.28), awful (0.28), stressful (0.27), stress (0.26), fabulous (0.26), felt (0.25), heart (0.24), lucky (0.24), cried (0.23), overwhelming (0.23), sleep (0.23), hours (0.22), scared (0.22), sick (0.22), therapy (0.22), am (0.22), myself (0.22), feels (0.22)
Adventurousness	Streets (0.28), city (0.27), century (0.25), sexual (0.24), industry (0.24), businesses (0.24), south (0.23), tour (0.23), Sean (0.23), global (0.22), diaper (0.22), immigration (0.22), countries (0.22), legal (0.22), poet (0.22), buildings (0.22), employment (0.22), west (0.21), little (0.21), al (0.21)
Intellect	Against (0.37), argument (0.35), knowledge (0.35), by (0.34), sense (0.34), political (0.34), models (0.34), belief (0.34), human (0.34), historical (0.33), greater (0.33), state (0.33), universe (0.33), philosophy (0.33), humans (0.33), beings (0.33), evidence (0.32), scientists (0.32), thank (0.32), leap (0.32)
Liberalism	Complicated (0.4), literature (0.37), particularly (0.37), prayers (0.36), giveaway (0.36), thankful (0.35), hubby (0.34), let (0.34), unlikely (0.34), less (0.33), complex (0.33), folk (0.33), terms (0.33), fucking (0.33), entirely (0.33), structure (0.33), cultural (0.33), liberal (0.33), university (0.32), bizarre (0.32)
Agreeableness trait	Top 20 words
Trust	Summer (0.31), afternoon (0.29), spent (0.27), exploring (0.27), fuck (0.25), finishing (0.25), early (0.24), evening (0.24), Reagan (0.24), visiting (0.24), harm (0.23), year (0.23), drugs (0.23), USA (0.23), spring (0.23), two (0.23), minute (0.23), excuse (0.23), amendment (0.23), planned (0.23)
Morality	UK (0.26), finish (0.25), gifts (0.24), nap (0.24), finished (0.24), laundry (0.24), popcorn (0.24), day (0.23), goodness (0.23), blessed (0.23), two (0.23), guardian (0.23), through (0.23), rest (0.23), gray (0.22), bin (0.22), folded (0.22), sexual (0.22), book (0.22), until (0.22)
Altruism	Idiot (0.24), hug (0.24), blast (0.23), chips (0.23), greeted (0.23), minutes (0.22), rest (0.22), times (0.22), cup (0.22), beach (0.22), solved (0.22), seconds (0.22), Olympic (0.22), stupid (0.22), following (0.21), dinner (0.21), participants (0.21), die (0.21), fabulous (0.21), sharing (0.21)
Cooperation	Fuck (0.3), unusual (0.3), asshole (0.28), spring (0.27), particular (0.26), porn (0.25), lake (0.25), paid (0.25), seemed (0.25), two (0.25), fucking (0.25), enemies (0.24), sexual (0.24), tree (0.24), four (0.24), adventure (0.24), determined (0.23), gay (0.23), occasionally (0.23), activity (0.23)
Modesty	Audience (0.25), increasingly (0.25), decades (0.25), doctor (0.24), recent (0.24), toys (0.24), cities (0.23), streets (0.22), infection (0.22), style (0.22), city (0.21), crowds (0.21), decade (0.21), Russian (0.21), box (0.21), involves (0.21),category (0.21), cherry (0.21), model (0.21)

(*continued on next page*)

Table 5 (*continued*).

Sympathy	Particular (0.26), since (0.24), strength (0.24), information (0.24), assured (0.24), anyways (0.23), require (0.23), providing (0.23), increased (0.22), courage (0.22), particularly (0.22), hoped (0.22), health (0.22), t (0.22), em (0.22), fascinating (0.22), conversation (0.22), ways (0.21), fewer (0.21), children (0.21)
Conscientiousness trait	Top 20 words
Self-efficacy	Fired (0.23), Roberts (0.22), rough (0.21), Hawaii (0.21)
Orderliness	Desperate (0.27), routine (0.26), tbsp (0.26), vegetables (0.25), garlic (0.24), temperature (0.24), carrots (0.23), melted (0.23), snack (0.22), salad (0.22), popcorn (0.22), ps (0.22), days (0.22), terror (0.22), jail (0.21), warm (0.21), enjoying (0.21), with (0.21), extreme (0.21), cheese (0.21)
Dutifulness	Rest (0.26), fuck (0.26), popcorn (0.24), hr (0.23), 14 (0.23), intelligent (0.23), 4 (0.22), deck (0.22), bang (0.22), pity (0.22), 5 (0.22), lots (0.21), stack (0.21), 8 (0.21), 2 (0.21), finished (0.21), determine (0.21), pathetic (0.21), visit (0.2), extreme (0.2)
Achievement striving	Stupid (0.29), idiot (0.26), religious (0.25), vain (0.25), decent (0.25), wallet (0.24), deny (0.24), rarely (0.24), bloody (0.23), protest (0.23), utter (0.23), contrary (0.22), shame (0.22), majority (0.22), soldiers (0.22), drunk (0.22), politically (0.22), democracy (0.22), fuck (0.22), entirely (0.21)
Self-discipline	Practical (0.26), ready (0.25), HR (0.23), rarely (0.23), boring (0.23), quality (0.23), overcome (0.23), mom's (0.23), characters (0.22), bay (0.22), 8 (0.22), it is (0.22), involve (0.21), until (0.21), completed (0.21), with (0.21),entirely (0.21), clever (0.21), Mexican (0.2), idea (0.2)
Cautiousness	Cheap (0.23), rest (0.23), recovery (0.22), pace (0.22), challenging (0.22), addition (0.22), swear (0.22), bar (0.22), enjoy (0.21), anxious (0.21), fuck (0.21), jokes (0.21), terrific (0.21), extent (0.2), paid (0.2)

step increases, which once again proves the rationality of our parameter sensitivity test.

5.2.5. Case study

To interpretively illustrate the effectiveness of introducing psycholinguistic knowledge into our method, we pick two typical case studies based on PMN w/o PD, PMN, and HPMN. We further visualize the attention layer (i.e. the readout function), as illustrated in Fig. 8. Red denotes the attentive weight of the corresponding word. A deeper color indicates that the model pays more attention to the associated word. This process interprets our approach in terms of how word-level knowledge impacts a document's personality traits understanding.

For example, in Fig. 8(a), it is observed that the PMN model can catch the personality word "damn", while the PMN w/o PD ignores it, and the HPMN can pay more attention to the words near "damn". In Fig. 8(b), we observe that the personality words "sooooooo" and "lucky" are well captured. This phenomenon indicates that the PMN can not only capture personality words but also obtain contextual information under personality awareness. Furthermore, combined with the constructed heterogeneous graph, our model can obtain the interactions between words and psycholinguistic knowledge. Compared with the PMN, the HPMN pays more attention to partial personality words. It seems that not all personality words have the same effect on the detection task because personality words are highly related to certain traits. That is, combined with explicit lexical and knowledge relations, the psycholinguistic knowledge of different word levels can be utilized, which can help our model to judge textual personality polarity.

To further validate the interpretability of our model, we evaluate the recurrence ratio between identified high-signal words and personality words for each personality trait. Specifically, when predicting each personality trait, we choose the words with the top 10 attention weights in the readout function and judge whether they are the corresponding personality words. Table 3 shows the average percentage of repeated words between the identified top 10 words and the personality words in the test sets of MyPersonality and Essays.

Overall, the ratios of personality words to be selected by the readout function are relatively high in the two datasets, especially in the Essays dataset. This demonstrates that our model is capable of capturing personality words that are highly correlated with traits. At the same time, our model is more sensitive to personality words when predicting HEU and EXT traits, while ARG is less pronounced. This suggests that personality words have different degrees of relevance for some personality dimensions. As we discussed in Section 5.2.1, this may be because people with NEU and EXT traits often use modal particles and adjectives that are more recognizable, while people with AGR traits have a more peaceful tone. In general, our proposed model seems to identify texts that imply personality. Moreover, from the verification perspective, our model provides support for verifying the rationality of the vocabulary hypothesis to a certain extent.

6. Conclusion and future research

In this paper, we present a novel personality detection model with lexical psycholinguistic knowledge guild, which not only achieves accurate personality detection results for social media texts but also enables us to explore the interpretability of personality traits via word knowledge. First, we summarize a personality dictionary containing 2043 words and learn personality-aware word embeddings to refine more accurate word vectors. Then, in combination with the background psychological knowledge, we construct a heterogeneous word graph for each user. Finally, upon constructed heterogeneous graph, we propose a heterogeneous personality message-passing model. Through the interactions among heterogeneous nodes, we fully express the personality traits contained in the user's language. Our model is validated on the MyPersonality, Essays, YouTube, and PAN datasets and can generate superior personality detection results. At the same time, our model provides support for lexical hypotheses in psycholinguistic research from a computational linguistic perspective.

The Big Five personality traits are not independent. In this study, we only predict each personality trait individually. In fact, some correlations are present between the different personality traits. In the future, we will design a neural model for joint personality detection tasks in subsequent research.

CRediT authorship contribution statement

Yangfu Zhu: Conceptualization, Methodology, Software, Writing – original draft. **Linmei Hu:** Methodology, Writing – review & editing. **Nianwen Ning:** Writing – review & editing. **Wei Zhang:** Writing – review & editing. **Bin Wu:** Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported by the NSFC-General Technology Basic Research Joint Funds, China under Grant (U1936220), the National Natural Science Foundation of China under Grant (61972047) and the National Key Research and Development Program of China (2018YFC0831500).

Appendix. Psycholinguistics empirical research

Some psychologists [20,21] investigated the relationship between personality and language use at the level of LIWC category and individual words. As an influential research in psychology, Tal Yarkoni reported the results of a large-scale analysis of personality and word use in a large sample of blogs. Tables 4 and 5 summarize the results and present the top correlations for each of the Big Five traits and 30 facets, respectively.

References

[1] S. Dhelim, N. Aung, H. Ning, Mining user interest based on personality-aware hybrid filtering in social networks, Knowl.-Based Syst. 206 (2020) 106227, http://dx.doi.org/10.1016/j.knosys.2020.106227.

[2] C. Yin, X. Zhang, L. Liu, Reposting negative information on microblogs: Do personality traits matter? Inf. Process. Manage. 57 (1) (2020) 102106, http://dx.doi.org/10.1016/j.ipm.2019.102106.

[3] T. Shen, J. Jia, Y. Li, Y. Ma, Y. Bu, H. Wang, B. Chen, T.-S. Chua, W. Hall, Peia: Personality and emotion integrated attentive model for music recommendation on social media platforms, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34, 2020, pp. 206–213.

[4] H. Wang, Y. Zuo, H. Li, J. Wu, Cross-domain recommendation with user personality, Knowl.-Based Syst. 213 (2021) 106664, http://dx.doi.org/10.1016/j.knosys.2020.106664.

[5] C. Xu, Z. Guan, W. Zhao, Q. Wu, M. Yan, L. Chen, Q. Miao, Recommendation by users' multimodal preferences for smart city applications, IEEE Trans. Ind. Inf. 17 (6) (2020) 4197–4205.

[6] A. Guo, J. Ma, S. Tan, G. Sun, From affect, behavior, and cognition to personality: an integrated personal character model for individual-like intelligent artifacts, World Wide Web 23 (2) (2020) 1217–1239, http://dx.doi.org/10.1007/s11280-019-00713-w.

[7] J.W. Pennebaker, M.E. Francis, R.J. Booth, Linguistic inquiry and word count: LIWC 2001, Mahway: Lawrence Erlbaum Assoc. 71 (2001) (2001) 1–22.

[8] M. Coltheart, The MRC psycholinguistic database, Q. J. Exp. Psychol. A 33 (4) (1981) 497–505, http://dx.doi.org/10.1080/14640748108400805.

[9] T. Tandera, D. Suhartono, R. Wongso, Y.L. Prasetio, et al., Personality prediction system from Facebook users, Procedia Comput. Sci. 116 (2017) 604–611, http://dx.doi.org/10.1016/j.procs.2017.10.016.

[10] P.-H. Arnoux, A. Xu, N. Boyette, J. Mahmud, R. Akkiraju, V. Sinha, 25 tweets to know you: A new model to predict personality with social media, in: Proceedings of the International AAAI Conference on Web and Social Media, Vol. 11, 2017, pp. 472–475.

[11] D. Xue, L. Wu, Z. Hong, S. Guo, L. Gao, Z. Wu, X. Zhong, J. Sun, Deep learning-based personality recognition from text posts of online social networks, Appl. Intell. 48 (11) (2018) 4232–4246, http://dx.doi.org/10.1007/s10489-018-1212-4.

[12] N. Majumder, S. Poria, A. Gelbukh, E. Cambria, Deep learning-based document modeling for personality detection from text, IEEE Intell. Syst. 32 (2) (2017) 74–79, http://dx.doi.org/10.1109/MIS.2017.23.

[13] X. Sun, B. Liu, J. Cao, J. Luo, X. Shen, Who am I? Personality detection based on deep learning for texts, in: 2018 IEEE International Conference on Communications (ICC), 2018, pp. 1–6, http://dx.doi.org/10.1109/ICC.2018.8422105.

[14] S. Poria, A. Gelbukh, B. Agarwal, E. Cambria, N. Howard, Common sense knowledge based personality recognition from text, in: Mexican International Conference on Artificial Intelligence, 2013, pp. 484–496, http://dx.doi.org/10.1007/978-3-642-45111-9_42.

[15] Z. Ren, Q. Shen, X. Diao, H. Xu, A sentiment-aware deep learning approach for personality detection from text, Inf. Process. Manage. 58 (3) (2021) 102532, http://dx.doi.org/10.1016/j.ipm.2021.102532.

[16] Y. Mehta, S. Fatehi, A. Kazameini, C. Stachl, E. Cambria, S. Eetemadi, Bottom-up and top-down: Predicting personality with psycholinguistic and language model features, in: 2020 IEEE International Conference on Data Mining (ICDM), 2020, pp. 1184–1189, http://dx.doi.org/10.1109/ICDM50108.2020.00146.

[17] S. Han, H. Huang, Y. Tang, Knowledge of words: An interpretable approach for personality recognition from social media, Knowl.-Based Syst. 194 (2020) 105550, http://dx.doi.org/10.1016/j.knosys.2020.105550.

[18] V. Kaushal, M. Patwardhan, Emerging trends in personality identification using online social networks—a literature survey, ACM Trans. Knowl. Discovery Data (TKDD) 12 (2018) 1–30, http://dx.doi.org/10.1145/3070645.

[19] J.M. Digman, Personality structure: Emergence of the five-factor model, Annu. Rev. Psychol. 41 (1990) 417–440, http://dx.doi.org/10.1146/annurev.psych.41.1.417.

[20] C.H. Lee, K. Kim, Y.S. Seo, C.K. Chung, The relations between personality and language use, J. Gen. Psychol. 134 (4) (2007) 405–413.

[21] T. Yarkoni, Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers, J. Res. Personal. 44 (3) (2010) 363–373, http://dx.doi.org/10.1016/j.jrp.2010.04.001.

[22] R.J. Craig, Assessing personality and mood with adjective check list methodology: A review, Int. J. Test. 5 (3) (2005) 177–196, http://dx.doi.org/10.1207/s15327574ijt0503_1.

[23] C.E. Izard, D.Z. Libero, P. Putnam, O.M. Haynes, Stability of emotion experiences and their relations to traits of personality, J. Personal. Soc. Psychol. 64 (5) (1993) 847.

[24] W. Youyou, M. Kosinski, D. Stillwell, Computer-based personality judgments are more accurate than those made by humans, Proc. Natl. Acad. Sci. 112 (4) (2015) 1036–1040, http://dx.doi.org/10.1073/pnas.1418680112.

[25] C. Stachl, Q. Au, R. Schoedel, S.D. Gosling, G.M. Harari, D. Buschek, S.T. Völkel, T. Schuwerk, M. Oldemeier, T. Ullmann, et al., Predicting personality from patterns of behavior collected with smartphones, Proc. Emy Sci. 117 (30) (2020) 17680–17687, http://dx.doi.org/10.1073/pnas.1920484117.

[26] F. Mairesse, M.A. Walker, M.R. Mehl, R.K. Moore, Using linguistic cues for the automatic recognition of personality in conversation and text, J. Artificial Intelligence Res. 30 (2007) 457–500, http://dx.doi.org/10.1613/jair.2349.

[27] J. Pennington, R. Socher, C.D. Manning, Glove: Global vectors for word representation, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014, pp. 1532–1543, http://dx.doi.org/10.3115/v1/d14-1162.

[28] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, 2018, http://dx.doi.org/10.18653/v1/n19-1423, arXiv preprint arXiv:1810.04805.

[29] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, 2017, arXiv preprint arXiv:1706.03762.

[30] V. Lynn, N. Balasubramanian, H.A. Schwartz, 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, 2020, pp. 5306–5316, http://dx.doi.org/10.18653/v1/2020.acl-main.472.

[31] X. Sun, B. Liu, Q. Meng, J. Cao, J. Luo, H. Yin, Group-level personality detection based on text generated networks, World Wide Web (2019) 1–20, http://dx.doi.org/10.1007/s11280-019-00729-2.

[32] Z. Guan, B. Wu, B. Wang, H. Liu, Personality2vec: Network representation learning for personality, in: 2020 IEEE Fifth International Conference on Data Science in Cyberspace (DSC), 2020, pp. 30–37, http://dx.doi.org/10.1109/DSC50466.2020.00013.

[33] G. Xue, M. Zhong, J. Li, J. Chen, C. Zhai, R. Kong, Dynamic network embedding survey, Neurocomputing 472 (2022) 212–223.

[34] G. Hu, G. Lu, Y. Zhao, FSS-GCN: A graph convolutional networks with fusion of semantic and structure for emotion cause analysis, Knowl.-Based Syst. 212 (2021) 106584, http://dx.doi.org/10.1016/j.knosys.2020.106584.

[35] T. Bian, X. Xiao, T. Xu, P. Zhao, W. Huang, Y. Rong, J. Huang, Rumor detection on social media with bi-directional graph convolutional networks, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34, 2020, pp. 549–556.

[36] X. Song, J. Li, Y. Tang, T. Zhao, Y. Chen, Z. Guan, Jkt: A joint graph convolutional network based deep knowledge tracing, Inform. Sci. 580 (2021) 510–523.

[37] Z. Wang, C.-H. Wu, Q.-B. Li, B. Yan, K.-F. Zheng, Encoding text information with graph convolutional networks for personality recognition, Appl. Sci. 10 (2020) 4081, http://dx.doi.org/10.3390/app10124081.

[38] L. Yao, C. Mao, Y. Luo, Graph convolutional networks for text classification, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33, 2019, pp. 7370–7377, http://dx.doi.org/10.1145/3437963.3441746.

[39] E. Cambria, S. Poria, D. Hazarika, K. Kwok, SenticNet 5: Discovering conceptual primitives for sentiment analysis by means of context embeddings, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 32, 2018, pp. 1795–1802.

[40] H. Ahmad, M.Z. Asghar, A.S. Khan, A. Habib, A systematic literature review of personality trait classification from textual content, Open Comput. Sci. 10 (2020) 175–193, http://dx.doi.org/10.1515/comp-2020-0188.

[41] P. Zhao, L. Hou, O. Wu, Modeling sentiment dependencies with graph convolutional networks for aspect-level sentiment classification, Knowl.-Based Syst. 193 (2020) 105443, http://dx.doi.org/10.1016/j.knosys.2019.105443.

[42] Y. Zhang, X. Yu, Z. Cui, S. Wu, Z. Wen, L. Wang, Every document owns its structure: Inductive text classification via graph neural networks, in: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 2020, pp. 334–339, http://dx.doi.org/10.18653/v1/2020.acl-main.31.

[43] Y. Li, R. Zemel, M. Brockschmidt, D. Tarlow, Gated graph sequence neural networks, in: Proceedings of ICLR'16, 2016.

[44] J.W. Pennebaker, L.A. King, Linguistic styles: language use as an individual difference, J. Personal. Soc. Psychol. 77 (6) (1999) 1296.