Data Augmented Graph Neural Networks for Personality Detection

Yangfu Zhu, Yue Xia, Meiling Li, Tingting Zhang, Bin Wu*

Beijing University of Posts and Telecommunications, Beijing, China
zhuyangfu,meilinglee,zhangtingting,wubin@bupt.edu.cn, 1216918224@qq.com

Abstract

Personality detection is a fundamental task for user psychology research. One of the biggest challenges in personality detection lies in the quantitative limitation of labeled data collected by completing the personality questionnaire, which is very time-consuming and labor-intensive. Most of the existing works are mainly devoted to learning the rich representations of posts based on labeled data. However, they still suffer from the inherent weakness of the amount limitation of labels, which potentially restricts the capability of the model to deal with unseen data. In this paper, we construct a heterogeneous personality graph for each labeled and unlabeled user and develop a novel psycholinguistic augmented graph neural network to detect personality in a semi-supervised manner, namely Semi-PerGCN. Specifically, our model first explores a supervised Personality Graph Neural Network (PGNN) to refine labeled user representation on the heterogeneous graph. For the remaining massive unlabeled users, we utilize the empirical psychological knowledge of the Linguistic Inquiry and Word Count (LIWC) lexicon for multi-view graph augmentation and perform unsupervised graph consistent constraints on the parameters shared PGNN. During the learning process of finite labeled users, noise-invariant learning on a large scale of unlabeled users is combined to enhance the generalization ability. Extensive experiments on three real-world datasets, Youtube, PAN2015, and MyPersonality demonstrate the effectiveness of our Semi-PerGCN in personality detection, especially in scenarios with limited labeled users.

Introduction

Personality is the overall characteristics and manifestations of an individual in terms of their psychology and behavior (Fang et al. 2023). Personality detection aims to identify the personality traits implied in social media posts that offering a deeper insight into human behavior (Nutescu and Mocanu 2023), emotional processes (Lian, Liu, and Tao 2022), and mental health (Zanwar et al. 2023). Besides, it can provide timely and objective support for downstream applications, such as human–computer interaction systems (Chien, Chen, and Chan 2022), virtual dialogue systems (Yang, Chen, and Narasimhan 2021), and recommendation systems (Yang et al. 2022; Shen et al. 2020).

Early researchers primarily combined two sources of psychological lexicon such as Linguistic Inquiry Word Count (LIWC) (Tausczik and Pennebaker 2010) and Medical Research Council (MRC) (Coltheart 1981) to manually find statistical word usage patterns for identifying personality traits from the text. With the blossoming of social media, users are posting massive content daily that reveals their psychological activities, providing new opportunity for automatically inferring traits. Subsequently, Deep Neural Networks (DNNs) were employed to learn meaningful representations of posts to detect personality (Kampman et al. 2018; Sun et al. 2018). However, understanding the traits behind the posts is non-trivial. Recently, a line of efforts focused on the structure of posts to dig deeper into the relationship between language and personality traits, including using hierarchical attention network (Lynn, Balasubramanian, and Schwartz 2020), constructing a heterogeneous tripartite graph with psycholinguistic information (Yang et al. 2021b), learning dynamic graph neural networks for post set (Yang et al. 2023a).

Despite considerable progress in personality detection, existing models still suffer from the inherent weakness of the amount limitation of labels. Modern trait theory (John, Robins, and Pervin 2010) tries to model the personality by several dimensions and construct a questionnaire to measure their ground-truth traits. As a famous personality indicator, the Big Five personality inventory typically includes 50 or more items to answer, which is very time-consuming and requires a lot of human resources. Furthermore, due to privacy concerns, people are less willing to share personal trait information on the Internet. Hence, insufficient trained DNNs may limit the inference of personality from posts. The question of how to accurately recognize personality with limited labeled data remains unresolved.

To address the above issues, we similarly start from the structure of user-generated documents and propose psycholinguistic data augmented graph neural network to detect personality in a semi-supervised manner, called Semi-PerGCN. Specifically, we construct a heterogeneous personality graph for each labeled and unlabeled user, which includes three kinds of nodes, *i.e.,* user nodes, word nodes, and LIWC category nodes. Then a personality graph convolution network with path-specific attention is employed to refine labeled user representation on a heterogeneous graph. For un-

*Corresponding author

labeled users, we utilize the empirical psychological knowledge of the LIWC lexicon for multi-view graph augmentation, and perform unsupervised graph consistent learning on the parameters shared personality graph neural network. By incorporating noise invariant learning on large-scale unlabeled data during the learning process for finite labeled data, our Semi-PerGCN is more generalized for unseen data. In summary, the contributions of this work can be summarized as follows:

- To the best of our knowledge, this is the first effort to utilize the inherent signals of massive unlabeled data to promote finite labeled personality detection, which provides a new perspective to alleviate the dilemma of training data for such data-hungry tasks.
- We propose a novel psycholinguistic augmented graph neural network, for which based on heterogeneous graphs constructed by labeled and unlabeled users, supervised personality detection and unsupervised graph consistency learning are co-trained to avoid noise and improve generalizability.
- We conduct extensive experiments to demonstrate the effectiveness of Semi-PerGCN on three representative datasets (*i.e.,* Youtube, PAN2015, and MyPersonality) for personality detection.

Related Work

Personality Detection

In the field of psychology, most personality research focused on the Big Five, which encompasses five broad dimensions that describe human personality and psychology in a common language (Digman 1990). Manual personality scale measures are the standard methodology today, but they are difficult to meet the demands of large-scale investigations. Recently, personality computing has attracted the attention of psychologists and computer scientists due to its wide range of application scenarios. The scope of related research extends from early linguistic feature analysis to current deep learning methods.

Early researchers utilized psycholinguistic statistics features such as LIWC (Tausczik and Pennebaker 2010), Mairesse (Mehta et al. 2020a), and MRC (Tausczik and Pennebaker 2010) to assess personality, as they believe that personality traits affect language use patterns. With the rapid development of deep learning, DNNs are applied to personality detection task and achieve great success, such as Convolutional Neural Networks (CNNs) (Kampman et al. 2018), Recurrent Neural Networks (RNNs) (Sun et al. 2018), and Transformer (Yang et al. 2021a). Benefiting from large-scale pre-trained language models, A line of pretraining fine-tuning paradigm are also explored on this tasks, such as fine-tuning the BERT (Jiang, Zhang, and Choi 2020) and personality-specific prompt-tuning (Wen et al. 2023). Furthermore, another line of approaches focuses on the structure of user-generated documents. Hierarchical structure model incrementally aggregates documents from the post level and the user level (Lynn, Balasubramanian, and Schwartz 2020). Subsequently, TrigNet (Yang et al. 2021b) considers that there is a psycholinguistic structure between posts and integrates the information in the LIWC dictionary and posts by message-passing based graph neural network. D-DGCN (Yang et al. 2023a) holds a different view that structure between user-generated posts is agnostic and utilizes dynamic graph neural network to automatically learns the structure between posts.

The above methods mainly focus on how to obtain a meaningful post representation of users and rarely pay attention to the limitation of labeled data, which may restrict the generalization ability of the model when meet unseen data.

Graph Structure Learning on Text

Although the text is usually modeled as serialized tokens in Natural Language Processing (NLP) field, there are a large number of tasks that can be better modeled using graphs. In recent years, different GNN models have been applied to NLP tasks with great success. In these studies, how to build graphs to better capture textual information had attracted extensive attention. TextGCN (Yao, Mao, and Luo 2019) constructs a global heterogeneous graph containing word and document nodes for text classification. Then, an independent graph is constructed in (Huang et al. 2019) to represent each document, which greatly reduces the model's memory requirements and dependence on the corpus. TextING (Zhang et al. 2020) further points out that in a document-level graph, words in different texts should not share the same representation but should be trained separately. This method is obviously more suitable for inductive learning. Returning to our task, the above subsection mentioned TrigNet and D-DGCN exactly proves the feasibility of GNNs in modeling at the document structure level for personality detection.

Unlike TrigNet and D-DGCN, we represent user information by building co-occurring connections of words in different posts, which enhances the semantic learning of words. More importantly, we leverage massive unlabeled data for unsupervised training to stretch the upper bound of supervised detection models.

Consistency Regularization

Our work is also closely related to consistency regularization, which is used to smooth the output distribution, which is beneficial to model performance, even if the input data changes slightly, the output of the model can basically remain unchanged (Yang et al. 2023b). The consistency constraints of existing methods include the results of two forward operations for each sample participating in training: current model prediction results and historical predictions results (Tarvainen and Valpola 2017), model prediction results and prediction results after adding adversarial noise (Araslanov and Roth 2021), and model prediction results and prediction results after data augmentation (Zhao and Yao 2022). These methods are widely used in semi-supervised classification tasks, proving the effectiveness of using consistent regularization terms to extract information from unlabeled data. Based on the above successful enlightenment, we propose a personality graph neural network with consistent regularization for personality detection and make full use of the psychological knowledge to construct the disturbance agreement item, as well as extract the implicit personality information from the unlabeled data.

Preliminaries

Personality detection task can be regarded as a multi-document multi-dimensional regression problem. Mathematically, given a set of posts $\mathrm{P} = \{p_1, p_2...p_n\}$ of a user x, where $\mathrm{p}_i = \{w_i^1, w_i^2...w_i^k\}$ is the i-th post with k words, our goal is to learn a representation mapping function $\mathcal{F} : P = \{p_1, p_2...p_n\} \rightarrow Y$ to score t-dimensional personality traits intensity $Y = \{\mathrm{y}_1,\mathrm{y}_2...\mathrm{y}_t\}$ for this user based on the posts.

In this paper, we model user-generated documents as heterogeneous graphs over three different types of nodes, including user nodes, word nodes, and LIWC nodes. For each user, the heterogeneous graph can be represented as $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \mathcal{V}_u \cup \mathcal{V}_w \cup \mathcal{V}_l$ is a set of nodes and $\mathcal{E}$ represents the edges between nodes, $\mathcal{V}_u$ represents the user nodes, $\mathcal{V}_w = \{w_1, w_2, ..., w_p\}$ denotes p words that appear in the posts and $\mathcal{V}_l = \{l_1, l_2, ..., l_q\}$ is the q categories selected from LIWC 2015 dictionary. For the edges between word-user and word-LIWC are constructed undirected, indicating that words belong to the user or LIWC category, while for word-word, we simply use sliding windows to find their co-occurrence relationship and connect them without direction. And then based on the heterogeneous graph, we propose a semi-supervised graph neural network (Semi-PerGCN) for personality detection.

Methodology

The architecture of our proposed Semi-PerGCN is shown in Figure 1, which consists of two components, *i.e.,* supervised graph neural networks for personality detection and unsupervised augmented graph consistency learning. The personality graph neural network aims to learn rich user representations on heterogeneous user graphs via precious traits scores while the consistency learning component is to utilize unlabeled data for learning a more generalized model on such data-hungry task.

Personality Graph Neural Networks

Upon each personality graph, we employ Graph Convolutional Networks (GCNs) (Kipf and Welling 2016) to refine user presentation from both the global structure and the specific psycholinguistic structure of the heterogeneous user graph respectively.

Specifically, each convolutional layer in GCNs can process first-order neighborhood information, and multi-level neighborhood information transmission can be realized by superimposing several convolutional layers. The propagation and transformation are as follows:

$$X_{k+1} = \sigma(A\mathbf{X_k}\mathbf{W_k}), \tag{1}$$

where $\hat{A} = D^{-\frac{1}{2}}(A + I)D^{-\frac{1}{2}}$ is the normalized symmetric adjacency matrix, $\mathbf{W_k} \in \mathbf{R^{n\times n}}$ is a weight matrix, σ is the activation function.

Formally, based on constructed personality graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{A}, \mathcal{X})$, where $\mathbf{X}$ represents the initial node embedding of three types of nodes. We use BERT to initialize the word nodes $\mathcal{V}_w$ and randomly initialize the user nodes $\mathcal{V}_u$ and LIWC nodes $\mathcal{V}_l$. The A is the adjacency matrix of the heterogeneous personality graph $\mathcal{G}$.

We first update updates all node representations with two-layer GCNs as follows:

$$\mathbf{X_1} = \sigma(\mathbf{A}\mathbf{X_0}\mathbf{W_0}), \tag{2}$$

$$\mathbf{H} = \sigma(\mathbf{A}\mathbf{X_1}\mathbf{W_1}), \tag{3}$$

After two layers of iterations, we obtain the user $\mathbf{H_u}$ and LIWC $\mathbf{H_l}$ node representations under the global structure.

To further highlight the psychological structure of user-generated documents, attentional mechanisms on the specific path of user-LIWC are leveraged to aggregate summarized psycholinguistic information. As shown in Figure 2, $\mathbf{H_u}$ is the representation of a user used as the attention query and $[\mathbf{H}_l^1, \mathbf{H}_l^2,..., \mathbf{H}_l^q]$ as the key, the attention mechanism can be described as follows:

$$\beta_i = \delta(\mathbf{W_z}[\mathbf{W_u}\mathbf{H_u}||\mathbf{W_l}\mathbf{H}_l^i]), \tag{4}$$

$$\alpha_i = \frac{exp(\beta_i)}{\sum_{i=1}^{p} exp(\beta_i)}, \tag{5}$$

$$\widehat{\mathbf{H}}_\mathbf{u} = tanh(\sum_{i=1}^{p} \beta_i \mathbf{W_v}\mathbf{H}_l^i) + \mathbf{H_u}, \tag{6}$$

where $\mathbf{W_z}$, $\mathbf{W_u}$ and $\mathbf{W_l}$ are learnable linear transformation matrices. δ is the $LeakyReLU$ activation function. α_i is the attention weight for $\mathbf{H}_l^i$. The $\hat{\mathbf{H}}_\mathbf{u}$ is the weighted combination of the LIWC node and the user node V_u itself. Finally, the refine representation $\hat{\mathbf{H}}_\mathbf{u}$ of each labeled user x are fed into the linear layer and a sigmoid layer for the supervised training personality assessment model.

$$y_d = p_\theta(y|x) = \mathrm{Sigmiod}\left(\mathbf{W}_d\hat{\mathbf{H}}_u + b_d\right), \tag{7}$$

where $\mathbf{W_d}$ is a trainable weight matrix of the detection component with an output dimension set to 5 in accordance with the Big Five personality traits.

Unsupervised Augmented Graph Consistency Learning

We assume that a good detection model should be robust to any small change in the input examples and have good generalization to unseen data. Hence, we make full use of the large-scale unlabeled users by heterogeneous graph augmentations and design unsupervised data consistency regularization in the model training. In a nutshell, consistency training methods simply regularize model predictions to be invariant to small noise applied to input examples.

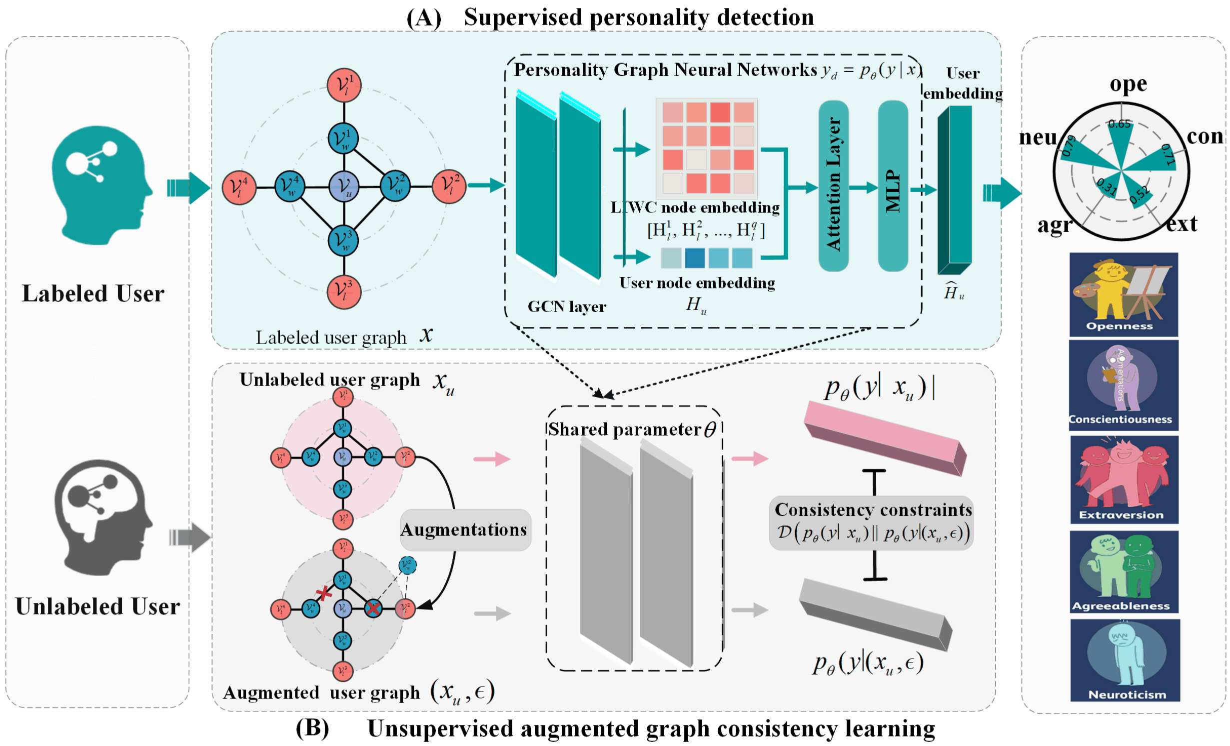


Figure 1: The overall architecture of our Semi-PerGCN, which consists of two components, (A) the supervised personality graph neural networks component aims to learn rich user representations on constructed heterogeneous user graphs via precious traits scores, and (B) the unsupervised augmented graph consistency learning component to utilize unlabeled data for helping to learn a more generalized model on such data-hungry task.

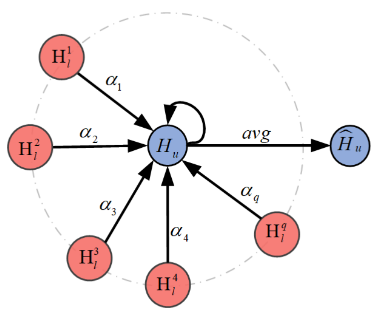


Figure 2: Learning user representation with LIWC information by attention mechanism.

Specifically, for any user x_u in unlabeled data, we compute the output distribution $p_\theta(y|x_u)$ by the supervised personality graph neural network in the above subsection, and then compute augmentation data with noise version $p_\theta(y|(x_u,\varepsilon))$ by injecting a small noise ε. Subsequently, the divergence metric between the two distributions $\mathcal{D}\left(p_\theta(y \mid x_u) || p_\theta(y \mid x_u,(x_u,\varepsilon))\right)$ is minimized by Cross Entrop (CE) loss:

$$\mathcal{L}_c = \mathrm{CE}\left(p_\theta(y \mid x_u) || p_\theta(y \mid x,(x_u,\varepsilon))\right), \tag{8}$$

where θ is a copy of the current parameters $\mathbf{W_k}$, $\mathbf{W_z}$, $\mathbf{W_u}$, $\mathbf{W_l}$, $\mathbf{W_d}$ of supervised detection modal indicating that the gradient is not propagated through θ. This approach ensures that the model is less affected by noise, resulting in a smoother response when there are variations in the input. Alternatively, by minimizing the consistency loss, the model gradually transfers label information from labeled examples to unlabeled ones.

However, how to construct a suitable sample pair $(x_u,(x_u,\varepsilon))$ is the key to the effect of consistency learning. As described in the classic unsupervised data augmentation method UDA (Zhang et al. 2021), more diverse and naturally more advanced data enhancements can lead to significant performance gains in supervised settings. Following this idea, we explored ways to augment data on this specific personality detection task. We are inspired by the lexical hypothesis of personality (Galton 1884), which posits that traits of personality are revealed by the descriptive vocabulary of human language. For a given user sample x_u, we randomly choose one of the following two operations to get (x_u,ε), as illustrated in Figure 3:

- LIWC Synonym Replacement (SR): Replace each psychology word with synonyms randomly. The psychology words belong to the same category in the LIWC dictionary as synonyms.
- Randomly Deletion (RD): Randomly removes some non-psychological words from user-generated documents.

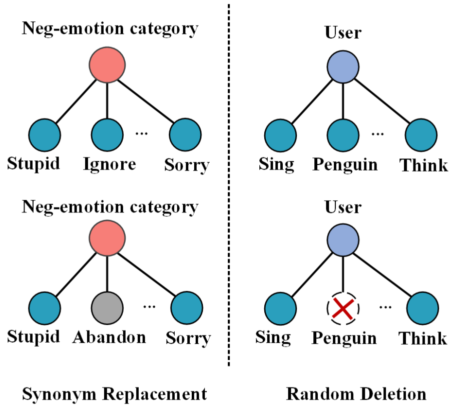


Figure 3: Two data enhancement methods on the personality graph, *i.e.,* word node replacement (left side) and word node deletion (right side).

In our personality graph neural network, words belonging to the same category are simultaneously connected to the same LIWC node. It can be regarded that performing synonym replacement in the text as adding noise to the attributes of word nodes in the graph. Similarly, randomly deleting words can be viewed as a slight disturbance to the structure of the graph. Therefore, we can regard the conversion method of x_u to (x_u, ε) as the data augmentation of the text, or as the addition of weak noise to the graph.

Objective Function

Following previous work, we minimize the objective function for the supervised detection process with the traditional MSE loss. The distance between the predicted personality score y_d and the ground-truth personality traits for the given data (x, y) minimize as:

$$\mathcal{L}_d = \mathrm{MSE}(\mathrm{y}_d, \mathrm{y}), \tag{9}$$

We jointly trained supervised detection and unsupervised detection under consistency constraints with labeled examples and large-scale unlabeled data. This auxiliary unsupervised detection task is included to help learn the invariant for the input noise, and the two tasks share the same parameters. The final optimization is to minimize the supervised MSE loss and the unsupervised consistency training loss, respectively.

$$\mathcal{L} = \mathcal{L}_d + \lambda \mathcal{L}_c, \tag{10}$$

where λ is a balance hyperparameter. In the actual training process, x represents the training data in the current mini-batch, and x_u is randomly selected from the unlabeled dataset.

Experiments

Experimental Settings

Datasets. Following previous studies (Mehta et al. 2020b), we conduct experiments on the Youtube Personality, PAN2015, and Mypersonality datasets with Big Five taxonomy. The number of users and posts of different datasets are shown in Table 1.

Datasets	User numbers	Post numbers
Youtube	404	3,205
PAN2015	294	27,344
MyPersonality	10,000	1,136,153

Table 1: Statistics of datasets.

- Youtube (Biel et al. 2013): consists of a collection of speech transcriptions with their Big Five personality scores which range from 1 to 7 [1]. The labels of this dataset are collected from the crowd-sourced annotation task. Annotators watch each video blog and then rated Big Five personality scores with a questionnaire.
- PAN2015 (Rangel Pardo et al. 2015): collected from the data science competition PAN2015 [2] and includes four languages datasets. We choose English data and their Big Five personality scores from -0.5 to 0.5.
- MyPersonality (Celli et al. 2013): collected from a Facebook application [3]. The data we use here is mainly collected from this work (Xue et al. 2018) and their Big Five personality scores from 0 to 5.

Evaluation Metrics. Following previous works, we choose MAE as our evaluation metric in each personality trait and use the average of each personality dimension MAE to measure their overall performance, formulated as:

$$\mathrm{MAE} = \frac{1}{\mathrm{N}} \sum_{\mathrm{i}=1}^{\mathrm{N}} |\mathrm{y_i} - \hat{\mathrm{y}}_\mathrm{i}|, \tag{11}$$

where N denotes the number of samples, $\hat{y}_i$ and y_i are the ground truth and the predicted personality trait scores of users.

Baselines. We compare our Semi-PerGCN with three groups of baseline methods: 1) Psychological vocabulary-based models; 2) Deep neural networks-based models 3) Document structure-based models, which can be categorized as follows:

Psychological vocabulary-based models: LIWC: LIWC features contain 89 features extracted by LIWC API. They are fed into the SVM model to predict the Big-Five personality traits. The reason why we choose this comparison method is that it mainly focuses on the lexical features of the text.

Deep neural networks-based models: TextCNN (Kampman et al. 2018): is a tri-modal architecture CNN to predict Big Five personality trait scores from video clips with different channels for audio, text, and video data. We use the CNN structure designed for text in their research. **TextCNN+LIWC** (Wei et al. 2017): concatenates

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

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

[3] http://mypersonality.org.

Datasets	Traits	Methods								
		LIWC	CNN	CNN+LIWC	2CLSTM	BERT	TrigNet	Attn	D-DGCN	Ours
Youtube	OPE	61.55	**60.54**	61.61	61.21	65.17	65.22	66.77	**64.23**	65.81
	CON	57.25	62.65	57.34	62.63	59.98	56.87	57.98	55.87	**55.01**
	EXT	90.84	86.83	88.68	86.62	86.28	85.02	84.91	**84.30**	85.24
	AGR	61.09	65.66	62.61	66.08	64.56	66.94	62.11	60.15	**56.34**
	NEU	64.18	62.39	64.71	66.94	61.81	61.88	61.77	61.46	**58.64**
	AVE	66.98	67.61	66.99	68.70	67.56	66.79	66.71	65.20	**64.21**
PAN2015	OPE	13.88	13.91	13.62	14.13	13.14	13.44	**12.78**	12.81	12.84
	CON	12.19	12.35	12.22	12.48	11.67	12.37	12.43	12.55	**12.17**
	EXT	13.12	13.22	13.25	13.34	13.21	13.33	12.83	12.67	**12.71**
	AGR	11.12	11.13	11.64	11.39	11.29	11.59	**11.12**	11.56	11.49
	NEU	18.62	18.59	18.51	18.24	19.93	16.72	18.12	18.34	**16.13**
	AVE	13.76	13.81	13.49	13.97	13.81	13.59	13.45	13.50	**12.98**
MyPersonality	OPE	54.49	55.91	55.90	60.37	55.75	55.35	53.68	52.23	**51.05**
	CON	54.97	58.49	57.33	61.41	57.24	56.48	55.78	55.57	**54.38**
	EXT	62.48	63.70	63.46	68.88	63.93	64.18	61.21	60.96	**60.42**
	AGR	56.08	56.80	56.84	60.65	56.67	56.86	55.80	55.37	**55.01**
	NEU	65.73	66.91	66.57	67.55	66.22	66.34	65.97	65.34	**64.80**
	AVE	58.75	60.36	60.02	63.77	59.96	59.84	58.49	57.90	**57.13**

Table 2: Performance comparison MAE (%) of our proposed Semi-PerGCN with baselines.

textual semantic features with the LIWC features for personality prediction. It combines deep neural networks and personality lexicon features for detection. **2CLSTM** (Sun et al. 2018): uses Bi-directional LSTM (Hochreiter and Schmidhuber 1997) and CNN to encode texts for detecting personality traits. **BERT** (Devlin et al. 2018): the fine-tuned BERT is firstly used to encode each post, and then mean pooling is performed over all posts to generate the user representation.

Document structure-based models: Attn (Lynn, Balasubramanian, and Schwartz 2020): is a hierarchical structure model which uses word-level attention to encode each post and another post-level attention to generate user representation. **TrigNet** (Yang et al. 2021b): is a psycholinguistic knowledge-based tripartite graph network that transmitting messages between neighboring parties in the posts graph by the specific flow. **D-DGCN** (Yang et al. 2023a): is a dynamic deep graph convolutional network that automatically learns the structure between user-generated posts.

Implementation Details. We use Pytorch to implement all the deep learning models on our three 2080Ti GPU cards. Empirically, we use a batch size of 16,16, and 64 for the labeled data and a batch size of 32, 32, and 112 for the unlabeled data in Youtube, PAN2015, and MyPersonality datasets respectively. Adam is utilized as the optimizer and the learning rate of our model is set to 0.0001, 0.0003, and 0.0003 in PAN2015, Youtube, and MyPersonality datasets respectively. The pre-trained language models BERT are employed to initialize the word node embeddings by the bert-base-cased (Devlin et al. 2018), and the dimensions of word nodes, LIWC nodes, and user nodes are set to 200. All the hyperparameters are tuned over the validation set to obtain the optimized results.

Overall Results

We compare our Semi-PerGCN to all baselines on the Youtube, PAN2015, and MyPersonality datasets. The MAE scores of different models are shown in Table 2. The major findings can be summarized as follows:

- The proposed Semi-PerGCN performs the best on all datasets. Especially, compared with the state-of-the-art baseline method GRU+Attn, Semi-PerGCN achieves 1.51%, 3.49%, and 1.32% improvements in average-MAE on Youtube, PAN2015, and MyPersonality datasets respectively. The results verify the effectiveness of our model in personality detection. We believe the reasons are two-fold: (1) Our model Semi-PerGCN uses a large amount of unlabeled data for consistency learning, which help to learn better user representations and reduces the risk of overfitting on a small training set. (2) Rich user representations are well captured by supervised graph learning on constructed heterogeneous user graphs.
- TrigNet performs slightly inferior to the GRU+Attn model, especially on the PAN2015 dataset, in which the redundancy of posts makes it difficult to obtain effective representations of the posts in TrigNet. Our model avoids this problem by directly modeling the user node. Dgcb is essentially the sub-optimal model on three datasets, which suggests that the upper bound of the supervised model may be harder to approximate. Additionally, the indicator on the PAN2015 dataset outperforms the other datasets due to the category imbalance in the PAN2015 dataset.
- We find the deep neural networks model 2CLSTMs performs the worst on all datasets, We think that the bidirectional LSTM may be not suitable for capturing long dependencies of text. The TextCNN performs slightly better than 2CLSTMs due to its capability to aggregate the

Datasets	Mean-MAE (%)		
	Sim-PerGCN	- Regularization	- Attention
OPE	65.81	64.85	66.13
CON	55.01	57.62	56.66
EXT	85.24	85.32	81.98
AGR	56.34	58.73	60.46
NEU	58.64	62.50	61.24
AVE	**64.21**	65.04 ($\downarrow$ 1.27%)	65.29 ($\downarrow$ 1.65%)

Table 3: Results of ablation study of Semi-PerGCN on Youtube, where "–" denotes the removal of a component, *i.e.,* "-Regularization" refers to remove the data augmentation component, and "-Attention" means remove the knowledge from LIWC.

associated textual information.

- TextCNN+LIWC performs better than TextCNN, which shows the effectiveness of psycholinguistic domain knowledge. Surprisingly, the model that only contains LIWC information performs better than CNN+LIWC. LIWC can directly capture psycholinguistic features and provide the most useful information for personality prediction. The poor performance of the CNN+LIWC combined model may be due to the conflict of features learned in CNN and LIWC.

Experimental Analysis

In this section, we analyze the usefulness of each component used in Semi-PerGCN and the effect of using different data augmentation strategies as well as the amount of unlabeled data.

Ablation Study. We conduct an ablation study for our Semi-PerGCN model on the Youtube dataset to investigate the effects of unlabeled data and psycholinguistic knowledge. As shown in Table 3, the performance of the model drops obviously after removing data augmentation or psycholinguistic knowledge information. The performance of the model drops 1.27% when we remove the consistency loss from Eq 10, which demonstrates the necessity of introducing external unlabeled data to make data augmentation for our task. Besides, the performance of the model drops 1.65% when we eliminate the impact of LIWC nodes and only use the user node representation after the interaction of the GCN layer. This shows the importance of psycholinguistic structure for personality detection.

Effect of Different Data Augmentation Methods. We investigate the effect of our model under different data augmentation strategies, including LIWC Synonym Replacement(SR), Randomly Deletion (RD), and both. The performances of our model under different data augmentation methods and different hyperparameters λ values settings in Eq 10 are shown in Figure 4. Compared to RD, SR is a more effective data enhancement method. This demonstrates that the LIWC lexicon, collected empirically by psychologists, is a suitable source of data enhancement on personality detection. Meanwhile, combining the two enhancements can play a positive role and achieve the best effect in our graph model. On the other hand, as the value of λ increases, the performance improves initially but eventually starts to decline beyond a certain point. This indicates that excessive unsupervised learning likewise weaken the model's generalization ability.

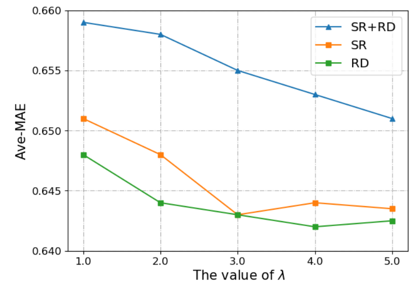
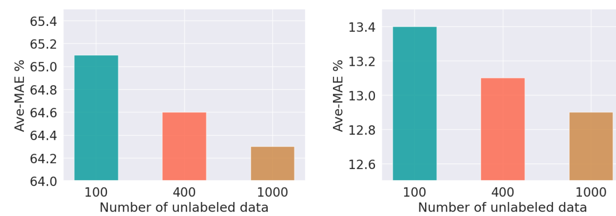


Figure 4: Performance curves f of different data augmentation methods and different hyperparameter λ values.



(a) Youtube (b) PAN2015

Figure 5: Results of adding different amounts of unlabeled data in Youtube and PAN2015 datasets.

Effect of Different Amounts of Unlabeled Data. We also explore the effect of different amounts of unlabeled data on the prediction results of the model. We extracted 100, 400, and 1000 unlabeled data for consistency training on the Youtube and PAN2015 datasets. As shown in Figure 5, performance improves as the amount of unlabeled data increases. It is worth noting that the performance improvement from 400 to 1000 is slower than that from 100 to 400, suggesting that there is an upper limit for the model to extract information from unlabeled data.

Conclusion

In this paper, we propose a semi-supervised graph neural network (Semi-PerGCN) for personality prediction. Semi-PerGCN expects to leverage the large amount of unlabeled data to help the model be insensitive to input noise. To this end, we construct a personality graph neural network that enhances generalization to unknown data by incorporating noise-invariant learning of large-scale unlabeled data during the learning process of limited labeled data. While the idea of data augmentation in a personality detection is validated, our implementation can be further improved by trying more metric learning methods such as contrastive learning to explore user representations.

Acknowledgments

This work is supported by the NSFC-General Technology Basic Research Joint Funds under Grant (U1936220), the National Natural Science Foundation of China under Grant (61972047, 62372060), and the BUPT Excellent Ph.D. Students Foundation (CX2022219).

References

Araslanov, N.; and Roth, S. 2021. Self-supervised augmentation consistency for adapting semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 15384–15394.

Biel, J.-I.; Tsiminaki, V.; Dines, J.; and Gatica-Perez, D. 2013. Hi YouTube! Personality impressions and verbal content in social video. In *Proceedings of the 15th ACM on International conference on multimodal interaction*, 119–126.

Celli, F.; Pianesi, F.; Stillwell, D.; and Kosinski, M. 2013. Workshop on computational personality recognition: Shared task. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 7.

Chien, S.-Y.; Chen, C.-L.; and Chan, Y.-C. 2022. The Influence of Personality Traits in Human-Humanoid Robot Interaction. *Proceedings of the Association for Information Science and Technology*, 59(1): 415–419.

Coltheart, M. 1981. The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology Section A*, 33(4): 497–505.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Digman, J. M. 1990. Personality structure: Emergence of the five-factor model. *Annual review of psychology*, 41(1): 417–440.

Fang, Q.; Giachanou, A.; Bagheri, A.; Boeschoten, L.; van Kesteren, E.-J.; Shafiee Kamalabad, M.; and Oberski, D. 2023. On Text-based Personality Computing: Challenges and Future Directions. In *Findings of the Association for Computational Linguistics: ACL 2023*, 10861–10879.

Galton, F. 1884. Measurement of character. Fortnightly Review, 36, 179-185. *Galton17936Fortnightly Review1884*.

Hochreiter, S.; and Schmidhuber, J. 1997. Long short-term memory. *Neural computation*, 9(8): 1735–1780.

Huang, L.; Ma, D.; Li, S.; Zhang, X.; and Wang, H. 2019. Text Level Graph Neural Network for Text Classification. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 3444–3450.

Jiang, H.; Zhang, X.; and Choi, J. D. 2020. Automatic Text-Based Personality Recognition on Monologues and Multiparty Dialogues Using Attentive Networks and Contextual Embeddings. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 13821–13822.

John, O. P.; Robins, R. W.; and Pervin, L. A. 2010. *Handbook of personality: Theory and research*. Guilford Press.

Kampman, O.; Barezi, E. J.; Bertero, D.; and Fung, P. 2018. Investigating Audio, Video, and Text Fusion Methods for End-to-End Automatic Personality Prediction. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, 606–611.

Kipf, T. N.; and Welling, M. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.

Lian, Z.; Liu, B.; and Tao, J. 2022. Pirnet: Personality-enhanced iterative refinement network for emotion recognition in conversation. *IEEE Transactions on Neural Networks and Learning Systems*, 1–12.

Lynn, V.; Balasubramanian, N.; and Schwartz, H. A. 2020. Hierarchical modeling for user personality prediction: The role of message-level attention. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5306–5316.

Mehta, Y.; Fatehi, S.; Kazameini, A.; Stachl, C.; Cambria, E.; and Eetemadi, S. 2020a. Bottom-up and top-down: Predicting personality with psycholinguistic and language model features. In *2020 IEEE International Conference on Data Mining (ICDM)*, 1184–1189. IEEE.

Mehta, Y.; Majumder, N.; Gelbukh, A.; and Cambria, E. 2020b. Recent trends in deep learning based personality detection. *Artificial Intelligence Review*.

Nutescu, C.-I.; and Mocanu, M. 2023. Creating personality model using genetic algorithms and behavioral psychology. In *15th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, 01–04.

Rangel Pardo, F. M.; Celli, F.; Rosso, P.; Potthast, M.; Stein, B.; and Daelemans, W. 2015. Overview of the 3rd Author Profiling Task at PAN 2015. In *CLEF 2015 evaluation labs and workshop working notes papers*, 1–8.

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

Sun, X.; Liu, B.; Cao, J.; Luo, J.; and Shen, X. 2018. Who am I? Personality detection based on deep learning for texts. In *2018 IEEE International Conference on Communications (ICC)*, 1–6. IEEE.

Tarvainen, A.; and Valpola, H. 2017. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. *Advances in neural information processing systems*, 30.

Tausczik, Y. R.; and 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.

Wei, H.; Zhang, F.; Yuan, N. J.; Cao, C.; Fu, H.; Xie, X.; Rui, Y.; and Ma, W.-Y. 2017. Beyond the words: Predicting user personality from heterogeneous information. In *Proceedings of the tenth ACM international conference on web search and data mining*, 305–314.

Wen, Z.; Cao, J.; Yang, Y.; Wang, H.; Yang, R.; and Liu, S. 2023. DesPrompt: Personality-descriptive prompt tuning for few-shot personality recognition. *Information Processing & Management*, 60(5): 103422.

Xue, D.; Wu, L.; Hong, Z.; Guo, S.; Gao, L.; Wu, Z.; Zhong, X.; and Sun, J. 2018. Deep learning-based personality recognition from text posts of online social networks. *Applied Intelligence*, 48(11): 4232–4246.

Yang, F.; Quan, X.; Yang, Y.; and Yu, J. 2021a. Multi-document transformer for personality detection. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, 14221–14229.

Yang, Q.; Nikolenko, S.; Huang, A.; and Farseev, A. 2022. Personality-Driven Social Multimedia Content Recommendation. In *Proceedings of the 30th ACM International Conference on Multimedia*, 7290–7299.

Yang, R.; Chen, J.; and Narasimhan, K. 2021. Improving Dialog Systems for Negotiation with Personality Modeling. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics*, 681–693.

Yang, T.; Deng, J.; Quan, X.; and Wang, Q. 2023a. Orders are unwanted: dynamic deep graph convolutional network for personality detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 13896–13904.

Yang, T.; Yang, F.; Ouyang, H.; and Quan, X. 2021b. Psycholinguistic Tripartite Graph Network for Personality Detection. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, 4229–4239.

Yang, X.; Song, Z.; King, I.; and Xu, Z. 2023b. A Survey on Deep Semi-Supervised Learning. *IEEE Transactions on Knowledge and Data Engineering*, 35(9): 8934–8954.

Yao, L.; Mao, C.; and Luo, Y. 2019. Graph convolutional networks for text classification. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, 7370–7377.

Zanwar, S.; Li, X.; Wiechmann, D.; Qiao, Y.; and Kerz, E. 2023. What to Fuse and How to Fuse: Exploring Emotion and Personality Fusion Strategies for Explainable Mental Disorder Detection. In *Findings of the Association for Computational Linguistics*, 8926–8940.

Zhang, Y.; Yu, X.; Cui, Z.; Wu, S.; Wen, Z.; and Wang, L. 2020. 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*, 334–339.

Zhang, Y.; Zhang, H.; Deng, B.; Li, S.; Jia, K.; and Zhang, L. 2021. Semi-supervised models are strong unsupervised domain adaptation learners. *arXiv preprint arXiv:2106.00417*.

Zhao, L.; and Yao, C. 2022. EICO: Improving Few-Shot Text Classification via Explicit and Implicit Consistency Regularization. In *Findings of the Association for Computational Linguistics: ACL 2022*, 3582–3587.