World Wide Web (2020) 23:1887–1906
https://doi.org/10.1007/s11280-019-00729-2

Group-level personality detection based on text generated networks

Xiangguo Sun[1] · Bo Liu[1] · Qing Meng[1] · Jiuxin Cao[2] · Junzhou Luo[1] · Hongzhi Yin[3]

Received: 25 April 2019 / Revised: 17 August 2019 / Accepted: 9 September 2019 /
Published online: 21 September 2019

Abstract

Personality analysis has been widely used in various social services such as mental healthcare, recommendation systems and so on because its natural explainability for AI applications in Web intelligence. With the penetration of Web2.0, traditional social researches have gradually turned to online social networks. However, for a long time, personality detection from online social texts has sunk into an embarrassing situation for the lack of large labeled datasets. Limited by supervised learning frameworks and small labeled datasets, prior works mainly detect one's personality in the individual perspective, which may not well meet the challenges of massive un-labeled data in the near future. In this paper, we present a first look into group-level personality detection and we use an unsupervised feature learning method instead of supervised methods used in most related works. We propose AdaWalk, a new and novel model of group-level personality detection by learning the influence from text generated networks. The model uses different kernels to evaluate how much a given node should decide its walk path locally or globally. The advantage of AdaWalk is three-folded: a) the model is an unsupervised feature learning method, which means it relies less on annotations. b) by traversing the network, we can capture the influence in the group level, thus the analysis of one's personality is not only based on individual records but also the information in groups. Therefore, AdaWalk can leverage small datasets more comprehensively. c) AdaWalk is scalable and can be easily transformed as distributed algorithms, which means it has more potential, compared with existing personality detection methods, to meet the massive data without annotations. We use AdaWalk to predict users' Big Five personality scores in **FIVE** heterogeneous personality datasets. Compared with more than **TEN** famous related methods, AdaWalk outperforms the others, meanwhile verifying the significance of the group perspective and unsupervised feature learning methods in the application of personality analysis. To make our experiment repeatable, AdaWalk and related datasets are available at https://xiangguosun.strikingly.com.

This article belongs to the Topical Collection: *Computational Social Science as the Ultimate Web Intelligence*
Guest Editors: Xiaohui Tao, Juan D. Velasquez, Jiming Liu, and Ning Zhong

✉ Xiangguo Sun
sunxiangguo@seu.edu.cn

Extended author information available on the last page of the article.

Keywords Personality psychology · Unsupervised features learning · Social media analysis · Online social networks

1 Introduction

In sociological research, Web2.0 has triggered a shift from traditional social study to big data mining for online social networks because of its big data volume and convenience. In the field of computational social science, personality [1, 14] refers to a synthesis of all the features which characterize one's pattern to other people or situations. It has been widely [21, 32, 40] acknowledged that personality is a significant explanation for one's various outward manifestations. Therefore it has broad applications [49] such as human-computer interaction [39, 54], mental healthcare [23], business analysis [29] and human resource management [7].

A generally accepted and influential metric in psychology for characterizing and measuring personality traits is the OCEAN Model (also known as the big five model) [13], which consists of five traits: **O**penness, **C**onscientiousness, **E**xtroversion, **A**greeableness, and **N**euroticism. Details of these five traits are introduced in Table 1.

A widely used method to obtain personality labels is the big five inventory.[1] Target users have to finish all of the questions in the inventory and their reports are then delivered to specialists for further analysis [8, 17, 18, 43]. Nevertheless, there are three main problems with this approach. First, the reliability of the results depends to a great extent on the participant's temporary psychological state, not his/her chronic, stable characteristics (an outgoing boy may be upset by an accident because of a breakup with his girlfriend when he implemented the questionnaire, for example). Second, the criteria of different experts may not be the same, leading to deviations from the practice [50]. Third, this method is high cost and time-consuming, and thus not a wise choice for human resource departments.

With the penetration of online social media in our lives, studying one's personality traits from their online social records instead of traditional questionnaires has been of great interest to the researchers due to its convenience, and efficiency. Many studies [3, 4, 6, 38] have suggested the feasibility of detecting users' personality traits from their generated social texts. Compared with traditional psychology researches, which are limited by the huge cost of artificial analysis, computer science has its unique advantage. Aiming at automatically analyzing data, researches from computer science have the potential to deal with the challenge of big data. Regrettably, there is no open published large labeled personality dataset because of the enormous costs for data annotation and the policy of privacy protection, let alone the labeled personality data with network topology. These situations have become severe obstacles for more thorough researches because most of the related works can only use supervised approaches on small datasets and research their samples independently, ignoring the mutual influence from the group, which is far from addressing the forthcoming challenges of massive data with limited annotations for personality analysis.

We have realized there exist three tough obstacles impeding further researches. The first one is how to introduce semi-supervised or unsupervised methods so that we can meet the challenges of massive sparsely labeled data in online social media. Once we solve this

[1] https://www.ocf.berkeley.edu/~johnlab/bfi.htm

Table 1 Overview of the OCEAN model

Traits	Profile
Openness	Those who score higher in the Openness inventory are more likely to accept new ideas and be more creative or unconventional.
Conscientiousness	The Conscientiousness dimension refers to how controlled and self-disciplined we are. Those who score higher in this trait are more likely to be plan oriented and well-organized. Those on the low end are more likely to be careless, easily distracted from tasks, and undependable.
Extroversion	Extraversion refers to the degree to which an individual is outgoing. Extroverts are sociable people who also tend to be energetic, optimistic, friendly, and assertive. They are energized and thrive off being around other people.
Agreeableness	People who score higher on this dimension are perceived as kind, empathetic and cooperative. They are helpful, trusting, and sympathetic. The opposite tend to be antagonistic and skeptical.
Neuroticism	Individuals scoring higher on neuroticism are more likely to experience worry, fear, anger, and frustration. They often respond to stressors badly and lack emotional stability

problem, another obstacle comes out that how to make the methods scalable so it can be used in large datasets. The third obstacle is how to analyze one's personality based on the influence of groups so that we can leverage limited datasets more comprehensively for better performance. In light of these problems, we here propose a new and novel method to detect one's personality from his generated texts. Our principle contributions are summarized as follows.

- In order to meet the challenges of massive data but with sparse personality labels in online social media, we are the first paper to introduce the network representation learning (NRL) method into the field of personality detection, which is an unsupervised feature learning method for personality detection and can be easily transformed as distributed algorithms.
- We are the first paper to predict one's personality traits based on the collaborative identification, which can significantly improve the performance. Experiment results on **EIGHT** heterogeneous datasets (five personality datasets and three non-personality datasets) compared with more than **TEN** related famous methods confirm our method's advantages and verify the significance of the group perspective and unsupervised methods in the application of personality analysis.

The meaning of our work is we are not only the first paper to push related researches into group-level, which can significantly improve the performance, but also the paper proposing new thinking on current dilemmas faced by academic peers. Through our frameworks, we have better adaptability on limited datasets, and meanwhile have more potential to meet the challenges of large online social datasets with sparse personality labels in the near future.

The rest of this paper is organized as follows. We present related works in Section 2 and the basic concept in Section 3. Our model will be introduced in Section 4. In Section 5, we conduct various evaluations in multi-class classification and regression prediction for personality perception. Finally, we conclude our work in Section 6.

2 Related works

Due to the broad potential applications [36, 45, 52], **personality detection** has gradually come into the sight of computer science researchers. Compared to traditional sociological projects, online social media provides a lower-cost way for personality capture. Although in the budding period, related works have achieved fruitful outcomes. Early researches mainly focused on the artificial feature design such as Mairesse [25], and the model used was relatively simple [24, 37]. However, these artificial features require long texts, while in online social networks, most users generated texts records (such as twitter, weibo and so on) are very short, which can not meet the requirement of these features. Recently, researchers start to get rid of artificial features by automatically supervised feature learning. Therefore deep learning methods have been evolutionally applied, from shallow neural networks [49] to more complex networks [26, 46]; from just text data to multi-modal data [19, 49]. Unfortunately, most datasets are too small to support the training process of these deep learning models.

Broadly, the data type used in this area mainly includes texts, images [16], videos [19] and likes [20]. Besides these types, network structures also deserve to be mentioned here because of the straightforward common, "Birds of a feather flock together." People with a closely related personality more often build connections on various social occasions [5]. As an effective network topology research method, researches on network representation learning (NRL) have become more and more active recently. It aims at transforming the total network or each node from tricky topology to tractable vector in low-dimensional space. For example, DeepWalk [35] transforms the network topology into a series of sequences by the random walk. Then they use word2vec method [28] to evaluate each node's representation. After that, many variations such as LINE [47] and node2vec [15] based on the random walk have emerged.

Recent years, network representation learning has been widely recognized as an effective feature learning method in various scenarios such as sentiment analysis [48], recommender systems [12], community detection [10] and so on. Despite the great potentialities, there is no prior work concerning predicting one's personality in this way. As previously mentioned, it is impractical to collect social network data with personality labels for each node. Consequently, most of the existing studies with the assumption of independent and identically distribution (iid) ignore mutual influence between individuals. In the following section, we display a phenomenon that the similarity between users' generated texts is correlated with their personality trait similarity. It suggests that although there are no open personality datasets with following networks or retweeting networks, we can still construct a network based on the similarity of their generated textual contents as an instantiation of peoples' various relationships in real life.

3 NRL for personality analysis

Due to the limitations in personality analysis which are mentioned above, it is quite natural to introduce the NRL model into this field. Reasons are threefold: Firstly, most NRL models belong to unsupervised learning methods, which means they rely less on data annotations. Secondly, the latest models are generally scalable for large size data, especially those based on the **random walk** because they are easy to be transformed as distributed algorithms.

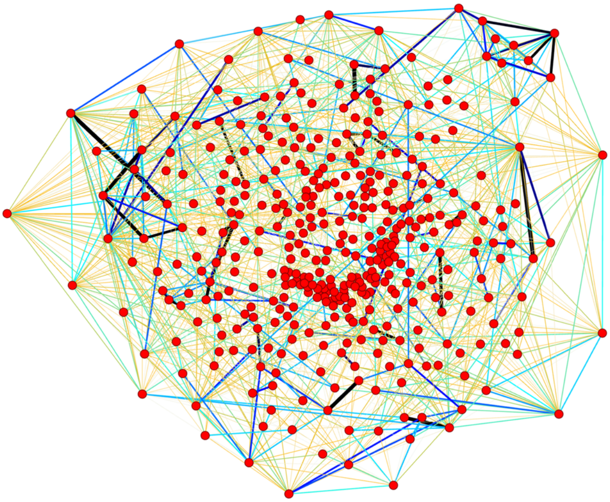


Figure 1 Texts Similarity w.r.t. Personality Similarity. This graph is generated from the Youtube dataset (see Section 5.2.1). The color and the width of edges describe the text similarity and personality similarity respectively. A pair of nodes connected by a thicker edge with deeper color are more similar both in personality and texts

Thirdly, NRL models can capture mutual influences in the dataset; thus they can perfectly match this application scenario in online social media.

To be specific, for each personality dataset, we first construct a complete graph $\mathcal{G}(\mathcal{V}, \mathcal{E}, \mathcal{W}, \mathcal{L})$. Here $\mathcal{V}$ is the user set. $\mathcal{E}$ is the edge set. For each edge $e(v_i, v_j)$, $v_i, v_j \in \mathcal{V}$, $w(v_i, v_j) \in \mathcal{W}$ measures the similarity between the texts generated by v_i and v_j. $\mathcal{L}$ is the set of users' Big Five personality[2] label. For each user v_i, his personality label is a vector with five entries, denoted by $\ell_i = [\ell_i^1, \ell_i^2, \ell_i^3, \ell_i^4, \ell_i^5]$. Then the similarity of labels from two users can be calculated by the cosine similarity of each pair of nodes.

Following above, we analyzed four personality datasets including Youtube, MyPersonality, PAN and OpenPsychometrics. Details of these datasets are given in Section 5.2.1. A sample network based on texts similarity in the Youtube dataset can be seen in Figure 1. We remove edges with very tiny weight ω. Edges' colors stand for the similarity of texts, while the width of edges describes the personality similarity. Two samples connected by a thicker edge with deeper color are more similar both in personality and texts, from which we can find that the correlation truly exists between personality and texts similarity. Results from other three datasets are very similar with Figure 1 and are not given here for brevity.

Problem definition Our problem can be treated as multi-classification or regression, depending on the values of labels. To be specific, for each user v_i, his personality label is a vector with five entries denoted by $\ell_i = \left[\ell_i^1, \ell_i^2, \ell_i^3, \ell_i^4, \ell_i^5\right]$. Each entry stands for the score

[2] A taxonomy for personality traits. It describe one's personality from five factors: Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism (or Stability).

in one personality trait. In most datasets, the entry value is continuous ([0,5] for example, stands for the confidence in one category). In other datasets, the entry is two-valued (for example, $\ell_i^j = 0$ means this user does not belong to trait j, $j = 1, 2, 3, 4, 5$ and vice versa). Given each user's feature vector as the input, we want to predict his personality traits scores across all five entries when the label values are continuous, or his trait categories when the label values are two-valued. In the rest of this section, we give some basic foundation of mathematics so that our model can be introduced more naturally.

Given a network $\mathcal{G}(\mathcal{V}, \mathcal{E}, \mathcal{W}, \mathcal{L})$, we want to learn a node embedding matrix $\Phi \in \mathbb{R}^{|\mathcal{V}|\times d}$. The learning process can be seen as the SkipGram method [28], a language model that maximizes the co-occurrence probability among the neighbors given a node (word). To be specific, if we get the neighbor nodes $N(v_j)$ of $v_j, v_j \in \mathcal{V}$, where $N(v_j) = \{v_{j-\omega}, \cdots, v_{j+\omega} \backslash v_j\}$ and ω is context size, then the following objective function will be optimized:

$$\max_{\Phi} \sum_{v \in \mathcal{V}} \log \Pr(N(v)|\Phi(v)) \tag{1}$$

where:

$$\Pr(N(v_j)|\Phi(v_j)) = \prod_{\substack{i=j-\omega \\ i \neq j}}^{j+\omega} \Pr(v_i|\Phi(v_j)) \tag{2}$$

$\Pr(v_j|\Phi(v_j)$ can be calculated by hierarchical softmax or negative sampling [28]. In the following section, we propose a sampling method based on the random walk. After that, we transform a network into a sentence-like corpus and then use the above optimization method to learn our node embeddings.

4 AdaWalk: adaptive walk for NRL

In this section, we will discuss the main components of our algorithm and give some methods to accelerate computing.

4.1 AdaWalk method

Given an on-going random walk path $< v_1, \cdots, v_t >$, node v_{t+1} can be assigned by the following probability:

$$\Pr(v_{t+1}|v_t, v_{t-1}) = \begin{cases} \frac{\alpha(v_{t-1}, v_t, v_{t+1}) w(v_t, v_{t+1})}{Z}, & \text{if } (v_t, v_{t+1}) \in \mathcal{E} \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

where Z denotes the normalizing constant, $w(v_t, v_{t+1})$ is the weight of edge (v_t, v_{t+1}). Inspired by the node2vec model [15], we design a biased walk strategy controlled by $\alpha(v_{t-1}, v_t, v_{t+1})$, which is defined as:

$$\alpha(v_{t-1}, v_t, v_{t+1}) = \begin{cases} \frac{1}{g_1} K(v_{t+1}) \cdot K(v_{t-1}), & \text{if } d_{v_{t-1}, v_{t+1}} = 0 \\ K(v_{t+1}) \cdot K(v_{t-1}), & \text{if } d_{v_{t-1}, v_{t+1}} = 1 \\ \frac{1}{g_2} K(v_{t+1}) \cdot (1 - K(v_{t-1})), & \text{if } d_{v_{t-1}, v_{t+1}} = 2 \end{cases} \tag{4}$$

The difference between AdaWalk and node2vec is that, for node2vec, the parameters g_1, and g_2 can be useful when the model deals with diverse datasets but not sufficient for fine-grained situations within the same dataset. For example, a node near a clique is more likely to be attracted to explore around the clique, while a node far from the clique is more likely to explore like DFS. This problem can be solved by our proposed method because $K(v)$ denotes the importance or attraction of node v to other nodes, which can be calculated by:

- **Degree**: let $d(v)$ be the degree of node v, we calculate $K(v)$ as a normalized degree value like:

$$K(v) = \frac{d(v)}{|\mathcal{V}| - 1}$$

- **Clustering Coefficient**: let $T(v)$ is the number of triangles through node v, and $d(v)$ is the degree of v, then $K(v)$ can be calculated as follows:

$$K(v) = \frac{2T(v)}{d(v)(d(v) - 1)}$$

- **Page Rank** [31] : Page Rank computes a ranking of the nodes in the graph based the structure of the incoming links.
  $K(v_i)$ then can be calculated as:

$$K(v_i) = \frac{1 - \alpha}{|\mathcal{V}|} + \alpha \sum_{v_j \in M(v_i)} \frac{K(v_j)}{d^{out}(v_j)}$$

  where $M(v_i)$ is the set of pages that link to v_i, $d^{out}(v_j)$ is the number of outbound edges of v_i and α is the hyperparameter. Note that the value should be normalized before used in our model.

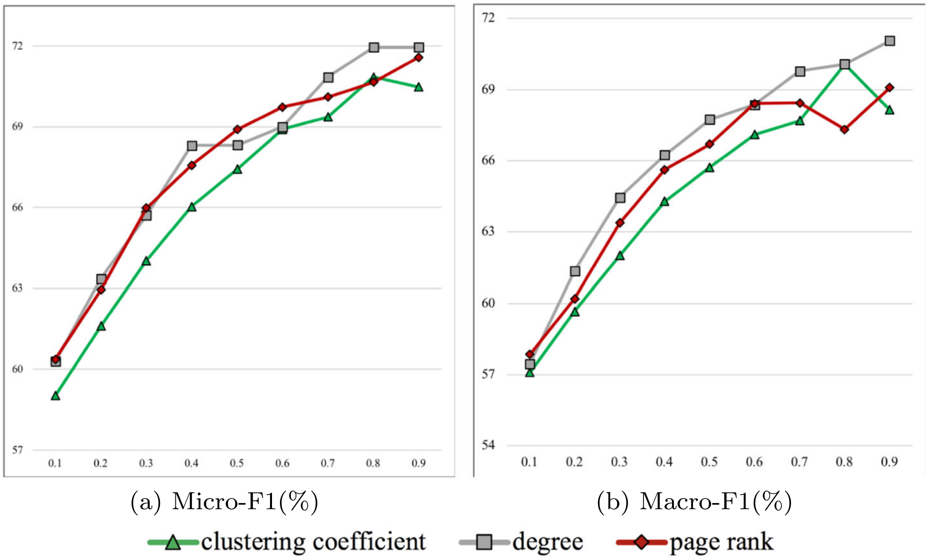
We compare the performance of different kernels in many related datasets, and Figure 2 is the results on Cora dataset (see Section 5.1.3), from which we can find that performances of different kernels are very close.

Intuitively, when selecting the consequent node for v_t, if the value of $K(v_{t-1})$ is big enough, the model is more likely to select those which are near to v_{t-1}. On the other hand, when the model faces some nodes in a similar situation, the candidate with a bigger value of $K(v_{t+1})$ is more likely to be chosen as the next hop. The pseudocode of AdaWalk is given in Algorithm 1, and our sampling strategy can be seen in Algorithm 1. There exist three phases of AdaWalk. Firstly, we preprocess the transition probabilities offline. Then the corpus is generated by a series of random walks with the complexity of $O(|\mathcal{E}|)$. Thirdly, optimization processing using SkipGram is executed sequentially.

4.2 Algorithm acceleration

We list some methods to accelerate our model:

- Firstly, some components can be calculated offline such as kernel K for each node, the transition probabilities, and each node's one-hop neighbors.



(a) Micro-F1(%)

(b) Macro-F1(%)

Figure 2 Different kernels on the Cora dataset. The horizontal axis stands for the percentage of training data and the vertical axis stands for Micro-F1 **a** and Macro-F1 **b** in the multi-classification task on the Cora dataset (see Section 5.1.3)

- Secondly, in the 10th line of Algorithm 2, we do not need to traverse all one-hop neighbors of node v_{curr}, and only a part of them is sufficient to make the model perform well. This strategy is especially useful when the network is a dense graph.
- Thirdly, the course of corpus generation (line 2-7 in Algorithm 1) can be executed parallelly because each iteration is independent of the others.
- Fourthly, according to [35], the frequency distribution of nodes in the corpus generated by random walks follows a power law, which means only a little number of nodes can affect the updating of Φ. Therefore the SkipGram Algorithm can be further optimized by the various parallel versions of the gradient descent method (asynchronous stochastic gradient descent with delay compensation [53], for example).

Algorithm 1 AdaWalk algorithm.

```
  Input: graph $\mathcal{G}(\mathcal{V}, \mathcal{E}, \mathcal{W})$, context size $\omega$, embedding size $d$, walk number per vertex
         $\gamma$, walk length $\ell$.
  Output: matrix of vertex representations $\Phi \in \mathbb{R}^{|V|\times d}$.
1 Initialization: Sample $\Phi$ from $\mathcal{U}^{|V|\times d}$, $walks = \emptyset$.
2 for $iter = 1$ to $\gamma$ do
3     for each vertex $u \in \mathcal{V}$ do
4         $walk$ =AdaRandomWalk($\mathcal{G}, u, \ell$)
5         Append $walk$ to $walks$
6     end
7 end
8 $\Phi$ = SkipGram($\omega, d, walks$)
9 return $\Phi$
```

Algorithm 2 AdaRandomWalk algorithm.

```
Input: graph $\mathcal{G}(\mathcal{V}, \mathcal{E}, \mathcal{W})$, start vertex $u$, walk length $\ell$, global tendency parameters:
    $g_1, g_2$.
Output: a random walk from graph $\mathcal{G}(\mathcal{V}, \mathcal{E}, \mathcal{W})$.
Initialization: $walk = [u]$,calculate kernel matrix $K$ for $V$.
$N_u = GetNeighbors(u, \mathcal{G})$
select a vertex $v \in N_u$ randomly.
Append $v$ to $walk$.
while $length(walk) < \ell$ do
    get the penultimate element in $walk$: $prior = walk[-2]$
    get the last element in $walk$: $cur = walk[-1]$
    $N_{cur} = GetNeighbors(cur, \mathcal{G})$
    $N_{prior} = GetNeighbors(prior, \mathcal{G})$
    for each vertex $x \in N_{cur}$ do
        if $x \in N_{prior}$ then
            $p_{cur,x} = \omega_{cur,x} * K_x * K_{prior}$
        else if $x == prior$ then
            $p_{cur,x} = \frac{1}{g_1} * \omega_{cur,x} * K_x * K_{prior}$
        else
            $p_{cur,x} = \frac{1}{g_2} * \omega_{cur,x} * K_x * (1 - K_{prior})$
        end
    end
    normalization: $p_{cur,x} = \frac{p_{cur,x}}{\sum_{t \in N_{cur}} p_{cur,t}}$
    select a vertex $x$ from $N_{cur}$ in probability $p_{cur,x}$, $x \in N_{cur}$.
    Append $x$ to $walk$.
end
return $walk$
```

5 Experiments

Studies of the personality detection mainly treat the task as multi-label classification or regression, depending on the label values of the personality datasets. As most personality datasets' labels are continuous numbers, it is more common to treat the task as regression. In this section, we want to measure our method from two aspects: (i) From Non-Personality datasets to Personality datasets. (ii) From Multi-Classification to Regression.

Firstly, as our method is an NRL model, it is quite necessary to compare it with other famous NRL methods in network learning. Through this, we can find that our method is better than the others and that is one of the reasons why we do not use other related NRL methods in our later personality detection. Therefore, in Section 5.1, we first compare our method with five famous NRL methods (Graph Factorization, DeepWalk, LINE, node2vec, HOPE) in the multi-classification task on three heterogeneous non-personality datasets and one personality dataset. We do this because the multi-classification task is one of the most important tasks in the NRL research field, and nearly conducted in every related NRL models. However, most personality datasets are not suitable for multi-classification because their label values are not discrete. At present, we only find one valid personality dataset with discrete label values, named stream-of-consciousness essays (SoCE). In order to make our assessment more convincing, we here supplement three more widely used datasets in most NRL models' evaluations (they are: BlogCatalog dataset, Cora dataset, and Wiki dataset). Secondly, we choose **EIGHT** famous personality detection methods as the baselines in the personality regression task in Section 5.2. We compare them with our AdaWalk model and report the root mean square error (RMSE) on **FOUR** diverse personality datasets.

STATEMENT In order to ensure the replicability to our experiments, we open our source code.[3] We have realized the challenge between privacy protection and the academic researches. Although all the datasets in this paper are obtained from open public sources, we still take very careful measures to achieve the best balance between academic needs and users' privacy protection, including but not limited to: a) removing user's sensitive information; b) acquiring the necessary license agreements before using related datasets; c) only for academic purpose.

5.1 Multi-class classification

5.1.1 Baselines

To validate the performance of our approach we compare it against the following baselines:

- Graph Factorization [2]: Graph Factorization is a distribute decomposition and inference framework for large-scale graphs.
- DeepWalk [35]: DeepWalk uses unbiased random walks to generate the corpus, then they use SkipGram to learn node embeddings.
- LINE [47]: LINE believes that a pair of nodes should be placed closely not only when they are connected (first-order proximity), but also when they share similar neighbors (second-order proximity). Therefore they use a breadth-first search strategy to generate context nodes and try to capture both the first-order proximity and the second-order proximity.
- node2vec [15]: On the basis of DeepWalk and LINE, node2vec discuss two search strategies: breadth-first search and depth-first search. Then they use a biased random walk in coarse-grained level to generate the corpus.
- HOPE [30]: HOPE incorporates Katz index to measure the proximity of nodes and preserves the asymmetric transitivity in the network. However, there exists a high time complexity of calculating high-order proximity measurements.

5.1.2 Parameter setup

Following above, we randomly sample a portion (70%) of the labeled nodes as training data. The rest of the nodes are used as the test set. We set the parameters as follows: the number of walks $\gamma = 30$, walk length $\ell = 80$, embedding size $d = 128$, context size $\omega = 20$. Specifically, for AdaWalk and node2vec, parameters g_1, g_2, p, q are selected from $[0.25, 0.5, 2]$. For LINE, Graph Factorization, and HOPE, we use the open source codes from OpenNE[4] and parameters are set in default. We repeat each test 10 times and report the

[3] The source code can be accessed only for the academic purpose. https://xiangguosun.strikingly.com

[4] An open-source toolkit for Network Embedding. https://github.com/thunlp/OpenNE

best performance in terms of both Macro-F 1 and Micro-F 1. For the downstream multi-class classifier, we use one-vs-rest logistic regression to return the most probable labels.

5.1.3 Datasets

We test our benchmarks on the following datasets:

- **SoCE** [33]: The stream-of-consciousness essays (SoCE) dataset contains 2, 467 persuasive anonymous essays tagged with the authors' personality traits: extroversion, neuroticism, agreeableness, conscientiousness, and openness. Each trait has a number choice from [0, 1], where 0 means the sample does not belong to this trait and vice versa.
- **BlogCatalog** [51]: it is an undirected network of social relationships from 10, 312 bloggers. The network contains 10, 312 nodes with 39 labels and 333, 983 edges.
- **Cora** [27]: Cora is a typical paper citation directed network. It contains 2, 708 nodes, 5, 429 edges and 7 labels.
- **Wiki** [44]: It contains 2, 405 Web pages from 19 categories and 17, 981 links between them.

5.1.4 Experimental results

As shown in Table 2, we evaluate the Micro-F1 score, and Macro-F1 score on BlogCatalog, Cora, Wikipedia, and SoCE with 70% labeled nodes. Numbers in bold represent the highest performance in each column. Note that we only sample 10% of one-hop neighbors in each random walk but achieve comparable performance with these models.

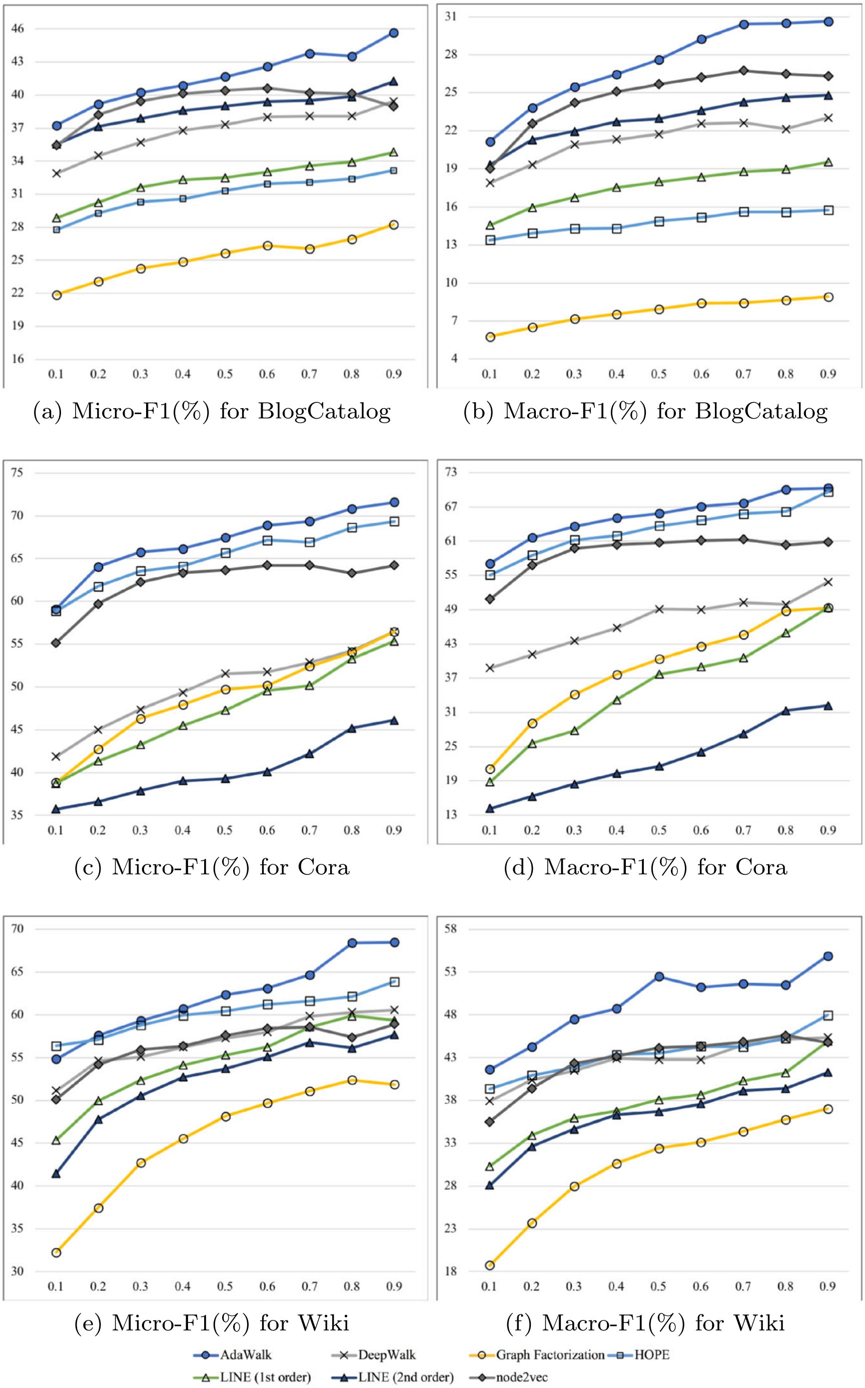
AdaWalk outperforms all baselines by at least 8% in BlogCatalog, 3% in Cora, and 7% in Wiki w.r.t. Micro-F1. Extraordinary, for personality dataset, SoCE, the Micro-F1 score of our method is up to **97.74%**. It also performs quite well in Macro-F1. We also note that node2vec generally performs better than DeepWalk, while our models perform better than node2vec, which suggests that it is meaningful to change the random walk from unbiased walk to biased walk with coarse-grained and fine-grained situations.

Additionally, we also change the labeled nodes percentage from 10% to 90% and compared with these baselines again. The results can be seen in Figure 3, from which we have observed that the performance of our models is very close to HOPE in Cora but beats most

Table 2 Multi-label classification results (70% labeled)

	Micro-F1 scores (%)				Macro-F1 scores (%)			
Algorithm	BlogCatalog	Cora	Wiki	SoCE	BlogCatalog	Cora	Wiki	SoCE
AdaWalk	**43.53**	**70.85**	**68.40**	**97.74**	**30.49**	**70.07**	51.50	**97.74**
node2vec	40.22	68.45	63.83	97.63	26.79	67.22	**53.98**	97.63
DeepWalk	38.09	54.24	60.29	97.64	22.13	49.92	45.18	97.63
Graph Factorization	26.92	54.06	52.39	61.46	8.66	48.83	35.79	61.20
HOPE	32.39	68.63	62.16	63.24	15.60	66.20	45.26	62.78
LINE (1st order)	33.92	53.32	59.88	62.54	18.96	44.90	41.25	62.47
LINE (2nd order)	39.86	45.20	56.13	64.05	24.63	31.31	39.39	63.96

BlogCatalog, Cora, and Wiki are non-personality datasets. SoCE is a personality dataset



(a) Micro-F1(%) for BlogCatalog

(b) Macro-F1(%) for BlogCatalog

(c) Micro-F1(%) for Cora

(d) Macro-F1(%) for Cora

(e) Micro-F1(%) for Wiki

(f) Macro-F1(%) for Wiki

Figure 3 Results w.r.t. labeled nodes percentage. Our model achieves the best performance even when the labeled nodes are reduced from 90% to 10%, which demonstrates the superiority when dealing with sparsely labeled cases

baselines in other datasets especially when the training set percentage down from 60% to 10%. **This is especially useful when we deal with massive data with sparse annotations.** The superiority of our model is even expanded when the training percentage up from 70% to 90%.

5.1.5 Parameter sensitivity

We compare our model with different one-hop neighbor sampling ratio r, walk length ℓ, and context size ω in the above multi-class classification task. Results are shown in Figure 4, from which we have the following observations:

- In order to accelerate computing, we sample each node's one-hop neighbors by the parameter r. From Figure 4a and b, we can find that both Micro-F1 and Macro-F1 in Cora and Wiki increase when r change from 0.1 to 0.2 but then both of them become stable, which suggests that there is no need to get the whole one-hop neighbors in each random walk. This strategy is especially necessary when the size of the dataset is large because compared with Cora and Wiki, performance in BlogCatalog w.r.t. r nearly does not fluctuate.
- We also research the performance when walk length $\ell > 30$. From Figure 4c and d, there is no drastic fluctuation when ℓ increases from 30 to 200. Besides, a larger dataset (BlogCatalog) performs more stable than smaller ones (Cora and Wiki).
- Finally, we also analyze the relationships between Micro/Macro-F1 and context size ω. From Figure 4e and f, the values of Micro/Macro-F1 scores have a little fluctuation when ω becomes larger. However, the performance differences are not that large in this case.

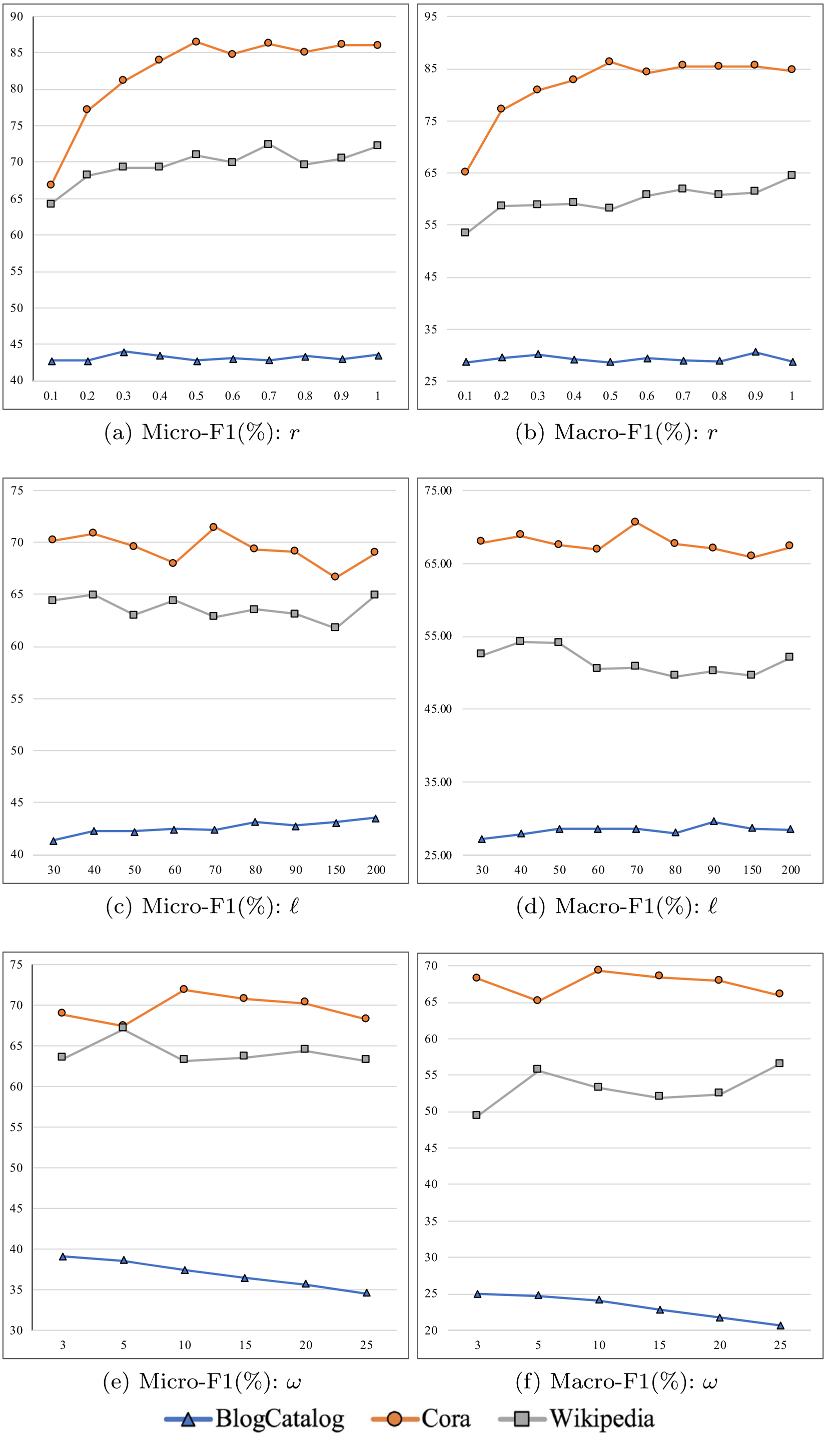
5.2 Regression prediction for personality

5.2.1 Datasets

Having compared our models with various baselines in multi-class classification, we now use AdaWalk to predict one's personality traits as the regression task. We extensively evaluate related models in four open published personality datasets, which are listed as follows:

- **Youtube Personality** [9]: This dataset consists of a collection of speech transcriptions with their Big Five personality impression scores (score value ranges from 1 to 7). We use it to generate a text similarity network which has 404 nodes and $81,406$ edges.
- **MyPersonality** [11]: This dataset comes from an open project on Facebook. It contains $9,900$ status updates from 250 users as well as their Big Five scores (range from 1 to 5). We translate this dataset into a similarity network, and it has 250 nodes with $31,125$ edges.
- **PAN** [42]: This dataset comes from a well-known data science competition, PAN2015, and it includes four languages (Dutch, English, Italian and Spanish) datasets. We select English data to construct the network and it contains 294 nodes with $43,070$ edges.
- **OpenPsychometrics**: This dataset comes from Open Source Psychometrics Project.[5] It contains 6 incomplete sentence responses, gender, age, and Big Five scores. We

[5] https://openpsychometrics.org



(a) Micro-F1(%): r

(b) Macro-F1(%): r

(c) Micro-F1(%): ℓ

(d) Macro-F1(%): ℓ

(e) Micro-F1(%): ω

(f) Macro-F1(%): ω

Figure 4 Results w.r.t. sampling ratio r, walk length ℓ, and context size ω

translate the dataset into a network based on the similarity of sentence responses. The network has 933 nodes with 434, 778 edges.

5.2.2 Baselines and parameter setup

Baselines are as follows:

a) by artificial features:

- mairesse. [25]: Mairesse expands the LIWC (linguistic inquiry and word count) dictionary which is developed by Pennebaker et al. [34] and makes the word dictionary become one of the most famous references when analyzing personality from texts.
- TF-IDF. [41]: TF-IDF is the product of the term frequency (TF) and inverse document frequency (IDF), which is one of the most basic and standard methods for text classification.

b) by supervised feature learning:

- 2CLSTMs [46]: In this work, they design a deep learning framework to process texts data so that they can detect the personality traits of the authors. The architecture of 2CLSTM can be divided into to parts. In the first part, they use bidirectional LSTMs concatenated with word vectors to encode sentence embeddings. In the second part, they use a group of CNNs to learn the sentence groups and generate the final feature vectors as the input of the downstream classifier.
- Kampman et al. [19]: They use deep learning method to detect one's personality from three channels (audio, text, and video). Specifically, for the text channels, they only use CNNs to finish the feature learning. Considering that we focus on text data, we only extract the text channel in our later experiment.
- Wei et al. [49]: They also use the CNN framework to process the text data, but before sending the final features learned by CNNs, they concatenate them with LIWC features.

c) by unsupervised feature learning:

- doc2vec [22]: This is a famous NLP model which is derived from Word2Vec and has been widely used in industry.
- DeepWalk: A NRL method mentioned in Section 5.1.1.
- node2vec: A NRL method mentioned in Section 5.1.1.

Given that above datasets are relatively smaller in the number of nodes, we split the training set and test set as $1 : 1$. We use SVR (support vector regression) to predict personality scores and RMSE (root mean square error) as the evaluation metric. In order to demonstrate that NRL methods are truly useful in this application scenario, we also calculate the results from random guess and take them as the worse case.

5.2.3 Experimental results

Results are shown in Table 3; we use EXT (Extraversion), AGR (Agreeableness), CON (Conscientiousness), EMO | STA | NEU (Emotion | Stability | Neuroticism), and OPN (Openness) to stand for the five traits in Big Five personality model. Note that in the Big Five personality model, researchers use different expressions to name the fourth dimension (EMO and NEU have the opposite value of STA).

Table 3 RMSE for personality prediction (50% labeled)

	MyPersonality					Youtube Personality				
Methods	EXT	NEU	AGR	CON	OPN	EXT	EMO	AGR	CON	OPN
AdaWalk	**0.7908**	0.7012	**0.6129**	**0.6642**	**0.4942**	**0.8897**	**0.6872**	**0.7743**	**0.6898**	0.6813
node2vec	0.8275	0.7285	0.6315	0.7140	**0.5081**	0.9088	0.7312	0.8113	0.7526	**0.6667**
DeepWalk	0.8270	**0.6830**	0.6463	0.7163	0.5097	0.9677	0.7854	0.9086	0.7648	0.7032
2CLSTMs	**0.7989**	0.7521	0.6441	0.7039	0.5476	0.9061	**0.7249**	0.8238	0.7474	0.6904
doc2vec	0.8388	**0.6770**	**0.6309**	0.7007	0.5404	**0.9023**	0.7385	0.8133	**0.6918**	0.6773
Kampman et al.	0.8413	0.7296	0.6531	0.7403	0.5533	0.9571	0.7446	0.8241	0.7453	**0.6536**
mairesse	0.8472	0.7475	0.6757	0.7665	0.5585	0.9300	0.7523	**0.7796**	0.7244	0.7276
TFIDF	0.8235	0.7353	0.6373	**0.6688**	0.5669	0.9623	0.7624	0.8276	0.7320	0.6952
Wei et al.	0.8793	0.7407	0.6417	0.7400	0.5571	0.9458	0.7688	0.8603	0.7170	0.7029
random guess	1.4311	1.4329	1.5054	1.4109	1.7104	4.2517	4.2750	4.1692	4.2927	4.2676

	PAN					OpenPsychometrics				
Methods	EXT	STA	AGR	CON	OPN	EXT	NEU	AGR	CON	OPN
AdaWalk	**0.1402**	0.2190	**0.1363**	0.1435	0.1489	**0.8715**	**0.8112**	**0.6935**	**0.7183**	0.6271
node2vec	0.1562	0.2245	0.1418	0.1392	0.1521	0.9204	0.8349	0.7325	0.7433	0.6257
DeepWalk	0.1559	0.2203	0.1460	0.1438	0.1556	**0.8818**	0.8265	0.7413	0.7562	0.6219
2CLSTMs	0.1613	0.2111	0.1461	0.1487	0.1445	0.9658	0.9001	0.7173	0.8011	0.6236
doc2vec	0.1507	0.2128	0.1476	**0.1376**	0.1424	0.8829	0.8320	0.7062	0.7623	0.6247
Kampman et al.	0.1563	**0.2076**	0.1494	0.1433	**0.1407**	0.9365	0.8519	0.7213	0.7617	**0.6067**
mairesse	**0.1424**	**0.1947**	0.1403	**0.1290**	**0.1371**	0.9256	0.8570	0.7134	0.7640	0.6294
TFIDF	0.1512	0.2139	**0.1395**	0.1425	0.1418	0.8886	**0.8196**	0.7092	**0.7402**	**0.6042**
Wei et al.	0.8351	0.7761	0.6855	0.7458	0.5210	0.9521	0.8677	**0.6972**	0.7584	0.6277
random guess	0.3755	0.4105	0.3619	0.3683	0.4013	3.0278	3.0693	3.1163	3.0320	3.0402

Entries in **bold** are the top two results

From Table 3 we can find that NRL models are truly useful in personality prediction. The values of RMSE are all lower than a random guess by 40%-96%. Furthermore, our model outperforms in most personality traits. Although significant advantages our model has achieved, we can still find that some artificial features such as mairesse still perform pretty well in some cases. For example, the results of mairesse on PAN dataset are very competitive but relatively poor on the other three datasets. This is because the text length in PAN dataset is relatively longer than the other datasets. Therefore the artificial features can perform well. However, as we previously mentioned before, in online social networks, most texts records are short, which means mairesse may not perform well in this scenario. We also notice that most supervised feature learning methods, especially those based on deep learning methods such as 2CLSTMs, Wei et al., and Kampman et al. are difficult to perform at full steam because of the limitation of datasets.

6 Limitations and conclusion

In this paper, we analyzed the feasibility of personality prediction from a group perspective. We introduce the NRL method to this field and extensively evaluate our model with other famous works. Results confirm the significance of the group perspective and unsupervised methods in the application of personality analysis.

However, this work also exists some limitations. For example, our text generated networks are constructed in advance based on the TF-IDF method. We do this because this is the simplest method to measure the similarity of two documents and can be conducted in all scenarios. Even though we've managed to build text networks, they are still the simulation of real-life relationships, not the real social networks such as following networks, retweeting networks and so on. However, to the best of our knowledge, there is no public social networks data with personality annotations, leaving our method be the most feasible choice.

Our work stands at a momentous crossing in the field of computational personality We are now living in an era of big data. There should have been much greater chance than ever before for researchers in social personality psychology to study one's personality in bigger data situations, especially in online social networks. However, for a long time, researchers have sunk into an embarrassing situation because they suffer from huge costs of personality annotation. Therefore, most open personality datasets are not big enough, resulting in related researches not adaptive for the forthcoming challenges from big data.

This contradictory situation has forced us to make a new thinking how to leverage the existing small labeled datasets more comprehensively, and meanwhile how to make our methods more scalable to deal with large-scale data in the forthcoming future. In light of above challenges faced by academic peers, we try to push related works to rely less on personality annotations, to leverage limited datasets more comprehensively, and to be more scalable for big data.

Acknowledgments This work is supported by National Key R&D Program of China 2017YFB1003000, National Natural Science Foundation of China under Grants No. 61972087, No. 61772133, No. 61472081, No. 61402104. Jiangsu Provincial Key Project BE2018706. Key Laboratory of Computer Network Technology of Jiangsu Province. Jiangsu Provincial Key Laboratory of Network and Information Security under Grants No. BM2003201, and Key Laboratory of Computer Network and Information Integration of Ministry of Education of China under Grants No. 93K-9.

References

1. Adelstein, J.S., Shehzad, Z., Mennes, M., DeYoung, C.G., Zuo, X.-N., Kelly, C., Margulies, D.S., Bloomfield, A., Gray, J.R., Castellanos, F.X., et al.: Personality is reflected in the brain's intrinsic functional architecture. PloS one **6**(11), e27633 (2011)
2. Ahmed, A., Shervashidze, N., Narayanamurthy, S., Josifovski, V., Smola, A.J.: Distributed large-scale natural graph factorization. In: Proceedings of the 22nd International Conference on World Wide Web, pp. 37–48. ACM (2013)
3. Amichai-Hamburger, Y., Vinitzky, G.: Social network use and personality. Comput. Hum. Behav. **26**(6), 1289–1295 (2010)
4. Andrew Schwartz, H., Eichstaedt, J.C., Kern, M.L., Dziurzynski, L., Ramones, S.M., Agrawal, M., Shah, A., Kosinski, M., Stillwell, D., Seligman, M.E.P., et al.: Personality, gender, and age in the language of social media: the open-vocabulary approach. PloS one **8**(9), e73791 (2013)
5. Bachrach, Y., Kosinski, M., Graepel, T., Kohli, P., Stillwell, D.: Personality and patterns of facebook usage. In: Proceedings of the 4th Annual ACM Web Science Conference, pp. 24–32. ACM (2012)
6. Bai, S., Gao, R., Zhu, T.: Determining personality traits from renren status usage behavior. In: Computational Visual Media, pp. 226–233. Springer (2012)
7. Barrick, M.R., Mount, M.K.: The big five personality dimensions and job performance: a meta-analysis. Person. Psychol. **44**(1), 1–26 (1991)
8. Benet-Martinez, V., John, O.P.: Los cinco grandes across cultures and ethnic groups: multitrait-multimethod analyses of the big five in Spanish and English. J. Person. Soc. Psychol. **75**(3), 729 (1998)
9. Biel, J.-I., Tsiminaki, V., Dines, J., Gatica-Perez, D.: Hi youtube!: personality impressions and verbal content in social video. In: Proceedings of the 15th ACM on International Conference on Multimodal Interaction, pp. 119–126. ACM (2013)
10. Cavallari, S., Zheng, V.W., Cai, H., Chang, K.C.-C., Cambria, E.: Learning community embedding with community detection and node embedding on graphs. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, pp. 377–386. ACM (2017)
11. Celli, F., Pianesi, F., Stillwell, D., Kosinski, M.: Workshop on computational personality recognition (shared task). In: Proceedings of the Workshop on Computational Personality Recognition (2013)
12. Chen, C.-M., Chien, P.-C., Lin, Y.-C., Tsai, M.-F., Yang, Y.-H.: Exploiting latent social listening representations for music recommendations. In: Proceedings of Ninth ACM International Conf. Recommender Syst. Poster (2015)
13. Costa, P.T., McCrae, R.R.: The revised neo personality inventory (neo-pi-r). The SAGE Handb. Person. Theory Assess. **2**, 179–198 (2008)
14. Deyoung, C.G.: Toward a theory of the big five. Psychol. Inq. **21**(1), 26–33 (2010)
15. Grover, A., Leskovec, J.: node2vec: scalable feature learning for networks. In: Proceedings of the 22nd ACM SIGKDD international Conference on Knowledge Discovery and Data Mining, pp. 855–864. ACM (2016)
16. Guntuku, S.C., Zhou, J.T., Roy, S., Lin, W., Tsang, I.W.: 'Who likes what and, why?' insights into modeling users' personality based on image 'likes'. IEEE Trans. Affect. Comput. **9**, 130–143 (2018)
17. John, O.P., Donahue, E.M., Kentle, R.L.: The big five inventory: versions 4a and 54, Institute of personality and social research. University of California, Berkeley (1991)
18. John, O.P., Naumann, L.P., Soto, C.J.: Paradigm shift to the integrative big five trait taxonomy. Handb. Person Theory Res. **3**, 114–158 (2008)
19. Kampman, O., Barezi, E.J., Bertero, D., Fung, P.: Investigating audio, video, and text fusion methods for end-to-end automatic personality prediction. In: ACL (2), pp. 606–611. Association for Computational Linguistics (2018)
20. Kosinski, M., Stillwell, D., Graepel, T.: Private traits and attributes are predictable from digital records of human behavior. Proc. Natl. Acad. Sci. **110**(15), 5802–5805 (2013)
21. Lambiotte, R., Kosinski, M.: Tracking the digital footprints of personality. Proc. IEEE **102**(12), 1934–1939 (2014)
22. Le, Q., Mikolov, T.: Distributed representations of sentences and documents. In: International Conference on Machine Learning, pp. 1188–1196 (2014)
23. Li, A., Zhang, F., Zhu, T.: Web use behaviors for identifying mental health status. In: International Conference on Brain and Health Informatics, pp. 348–358. Springer (2013)
24. Liu, X., Zhu, T.: Deep learning for constructing microblog behavior representation to identify social media user's personality. PeerJ Comput. Sci. **e81**, 2 (2016)

25. Mairesse, F., Walker, M.A., Mehl, M.R., Moore, R.K.: Using linguistic cues for the automatic recognition of personality in conversation and text. J. Artif. Intell. Res. **30**, 457–500 (2007)
26. Majumder, N., Poria, S., Gelbukh, A., Cambria, E.: Deep learning-based document modeling for personality detection from text. IEEE Intell. Syst. **32**(2), 74–79 (2017)
27. McCallum, A.K., Nigam, K., Rennie, J., Seymore, K.: Automating the construction of internet portals with machine learning. Inf. Retr. **3**(2), 127–163 (2000)
28. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Advances in Neural Information Processing Systems, pp. 3111–3119 (2013)
29. Nov, O., Ye, C.: Personality and technology acceptance: personal innovativeness in it, openness and resistance to change. In: Hawaii International Conference on System Sciences, Proceedings of the 41st Annual, pp. 448–448. IEEE (2008)
30. Ou, M., Cui, P., Pei, J., Zhang, Z., Zhu, W.: Asymmetric transitivity preserving graph embedding. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1105–1114. ACM (2016)
31. Page, L., Brin, S., Motwani, R., Winograd, T.: The pagerank citation ranking: bringing order to the Web. Technical report, Stanford InfoLab (1999)
32. Park, G., Andrew Schwartz, H., Eichstaedt, J.C., Kern, M.L., Kosinski, M., Stillwell, D., Ungar, L.H., Seligman, M.E.P.: Automatic personality assessment through social media language. J. Pers. Soc. Psychol. **108**(6), 934 (2015)
33. Pennebaker, J.W., King, L.A.: Linguistic styles: language use as an individual difference. J. Person. Soc. Psychol. **77**(6), 1296 (1999)
34. Pennebaker, J.W., Francis, M.E., Booth, R.J.: Linguistic inquiry and word count: Liwc 2001. Mahway: Lawrence Erlbaum Assoc. **71**(2001), 2001 (2001)
35. Perozzi, B., Al-Rfou, R., Skiena, S.: Deepwalk: online learning of social representations. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 701–710. ACM (2014)
36. Qian, Q., Huang, M., Zhao, H., Xu, J., Zhu, X.: Assigning personality/profile to a chatting machine for coherent conversation generation. In: IJCAI, pp. 4279–4285 (2018)
37. Qiu, L., Lin, H., Ramsay, J., Yang, F.: You are what you tweet: personality expression and perception on twitter. J. Res. Pers. **46**(6), 710–718 (2012)
38. Qiu, L., Lu, J., Ramsay, J., Yang, S., Qu, W., Zhu, T.: Personality expression in Chinese language use. International Journal of Psychology (2016)
39. Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C., Crandall, J.W., Christakis, N.A., Couzin, I.D., Jackson, M.O., et al.: Machine behaviour. Nature **568**(7753), 477 (2019)
40. Rai, T.S.: High replicability in personality psychology. Science **364**(6438), 348–348 (2019)
41. Rajaraman, A., Ullman, J.D.: Mining of massive datasets. Cambridge University Press, Cambridge (2011)
42. Rangel, F., Celli, F., Rosso, P., Potthast, M., Stein, B., Walter, D.: Overview of the 3rd author profiling task at PAN 2015. In: Cappellato, L., Ferro, N., Jones, G., San Juan, E. (eds.) CLEF Evaluation Labs and Workshop – Working Notes Papers, 8-11 September, Toulouse, France. CEUR-WS.org (2015)
43. Selfhout, M., Burk, W., Branje, S., Denissen, J., Van Aken, M., Meeus, W.: Emerging late adolescent friendship networks and big five personality traits: a social network approach. J. Person. **78**(2), 509–538 (2010)
44. Sen, P., Namata, G., Bilgic, M., Getoor, L., Galligher, B., Eliassi-Rad, T.: Collective classification in network data. AI Mag. **29**(3), 93 (2008)
45. Sun, G., Shen, J.: Facilitating social collaboration in mobile cloud-based learning: a teamwork as a service (taas) approach. IEEE Trans. Learn. Technol. **7**(3), 207–220 (2014)
46. Sun, X., Liu, B., Cao, J., Luo, J., Shen, X.: Who am I? personality detection based on deep learning for texts. In: ICC, pp. 1–6. IEEE (2018)
47. Tang, J., Qu, M., Wang, M., Zhang, M., Yan, J., Mei, Q.: Line: Large-scale information network embedding. In: Proceedings of the 24th International Conference on World Wide Web, pp. 1067–1077. International World Wide Web Conferences Steering Committee (2015)
48. Wang, H., Zhang, F., Hou, M., Xie, X., Guo, M., Qi, L.: Shine: signed heterogeneous information network embedding for sentiment link prediction. In: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, pp. 592–600. ACM (2018)
49. Wei, H., Zhang, F., Yuan, N.J., Cao, C., Fu, H., Xie, X., Rui, Y., Ma, W.-Y.: Beyond the words predicting user personality from heterogeneous information. In: Proceedings of the Tenth ACM International Conference on Web Search and Data Mining, pp. 305–314. ACM (2017)

50. Wu, Y., Kosinski, M., Stillwell, D.: Computer-based personality judgments are more accurate than those made by humans. Proc. Natl. Acad. Sci. **112**(4), 1036–1040 (2015)
51. Zafarani, R., Liu, H.: Social computing data repository at asu (2009)
52. Zhao, S., Ding, G., Han, J., Gao, Y.: Personality-aware personalized emotion recognition from physiological signals. In: IJCAI, pp. 1660–1667 (2018)
53. Zheng, S., Qi, M., Wang, T., Chen, W., Yu, N., Ma, Z.-M., Liu, T.-Y.: Asynchronous stochastic gradient descent with delay compensation. In: International Conference on Machine Learning, pp. 4120–4129 (2017)
54. Zibrek, K., Kokkinara, E., McDonnell, R.: The effect of realistic appearance of virtual characters in immersive environments - does the character's personality play a role? IEEE Trans. Vis. Comput. Graph. **24**(4), 1681–1690 (2018)

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Xiangguo Sun[1] · Bo Liu[1] · Qing Meng[1] · Jiuxin Cao[2] · Junzhou Luo[1] · Hongzhi Yin[3]

Bo Liu
bliu@seu.edu.cn

Qing Meng
qmeng@seu.edu.cn

Jiuxin Cao
jx.cao@seu.edu.cn

Junzhou Luo
jluo@seu.edu.cn

Hongzhi Yin
h.yin1@uq.edu.au

[1] School of Computer Science and Engineering, Southeast Univerisity, NanJing, 211189, China

[2] School of Cyber Science and Engineering, Southeast Univerisity, NanJing, 211189, China

[3] School of Information Technology and Electrical Engineering, The University of Queensland, Queensland, Australia