Information Processing and Management 58 (2021) 102532

Contents lists available at ScienceDirect

Information Processing and Management

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

A sentiment-aware deep learning approach for personality detection from text

Zhancheng Ren [a], Qiang Shen [b], Xiaolei Diao [c], Hao Xu [a,b,*]

[a] School of Artificial Intelligence, Jilin University, Changchun 130012, China
[b] College of Computer Science and Technology, Jilin University, Changchun 130012, China
[c] Department of Information Engineering and Computer Science, University of Trento, 38123, Italy

A R T I C L E I N F O

Keywords:
Personality detection
Deep learning
BERT
Multi-label classification

A B S T R A C T

Personality detection based on user-generated text content analysis has a significant impact on information science, for instance, information seeking. Existing deep learning-based approaches, however, have two major limitations. Firstly, they extract only keywords for personality detection and lack the analysis of sentiment information and psycholinguistic features. Secondly, the information about the context and polysemous words are ignored. To tackle these problems, we propose a novel multi-label personality detection model based on neural networks, which combines emotional and semantic features. Specifically, we leverage Bidirectional Encoder Representation from Transformers (BERT) to generate sentence-level embedding for text semantic extraction. In addition, a sentiment dictionary is used for text sentiment analysis in order to consider sentiment information. Finally, we input the above semantic information and emotional information into the neural network to construct an automatic personality detection model. The performance of the model has been evaluated on two public personality datasets. The experiments show that we obtain average accuracy improvements of 6.91% and 6.04% on the Myers-Briggs Type Indicator (MBTI) and Big Five datasets, respectively, compared with the state-of-the-art techniques.

1. Introduction

1.1. Background

Personality traits are described as fairly fixed characteristics of an individual, which indicate the individual's preferences and may influence the individual's decisions, and are incorporated into the methodology of information seeking, network security, clinical psychology, economics, and political theory, etc. (Al-Samarraie, Eldenfria & Dawoud, 2017; Russell, Weems, Ahmed & Richard, 2017). Human personality traits have been shown to influence some fundamental behaviors in life such as knowledge processing, emotion regulation, and information seeking (Taramigkou, Apostolou & Mentzas, 2018). Understanding the personality characteristics of users through social media text analysis can be considered as a task of information classification or information processing. An automated personality classification system can effectively help professional psychologists make judgments and decisions quickly, and help companies select more suitable candidates for their positions.

* Corresponding author.
E-mail address: xuhao@jlu.edu.cn (H. Xu).

https://doi.org/10.1016/j.ipm.2021.102532
Received 12 September 2020; Received in revised form 10 January 2021; Accepted 22 January 2021
Available online 2 February 2021

Modern trait theory (John, Robins & Pervin, 2010) tries to model a personality by setting several classification dimensions and constructing a questionnaire to measure them (Matthews, Deary & Whiteman, 1998). An example is the famous Big Five personality (Digman, 1990), including openness to experience, conscientiousness, extraversion, neuroticism, and agreeableness. The Myers-Briggs Type Indicator (MBTI) (Briggs Myers & Kirby, 2000) divides personality traits into four binary dimensions: extraversion versus introversion, sensing versus intuition, thinking versus feeling, and judging versus perceiving. However, traditional questionnaires have low timeliness, high cost, and difficulty adapting to the times, which led to the emergence of modern personality trait detection methods. In the past few decades, personality traits have been shown to be closely associated with users' social media behaviors (Azucar, Marengo & Settanni, 2018). Many algorithms use social media information to predict personality traits. There are mainly two types of personality trait detection methods. The first quantifies the text using the frequency of words and word categories and combines traditional machine learning methods (Support Vector Machine and Naive Bayes) for personality analysis. The representative method is Linguistic Inquiry and Word Count (LIWC) (Pennebaker, Francis & Booth, 2001). The second extracts textual information using deep neural networks for personality trait detection. The representative methods include the Convolutional Neural Network (CNN) (Fukushima, Miyake & Ito, 1988), the Recurrent Neural Network (RNN) (Elman, 1990).

Traditional machine learning and deep learning techniques can be effective at detecting the personality traits of users, and these methods prove that language-based models can accurately and efficiently identify the personality traits of users. However, there are still some challenges: 1) To deeply achieve an understanding of text semantics, most existing models for encoding text are based on static representations of word vectors, for example, Word2Vec (Mikolov, Chen, Corrado & Dean, 2013) and GloVe (Pennington, Socher & Manning, 2014), which cannot solve problems such as the polysemy of a word. Boxman-Shabtai et al. demonstrated that polysemy of a word in a text could lead to some ambiguous personality traits (Boxman-Shabtai & Shifman, 2014). 2) A challenging problem which arises in this domain is that current personality detection methods are far from being optimal. Therefore, the purpose of this paper is to answer the following question: How can automated personality detection from online open-source text be more accurately performed? This mainly includes the following points: 1) The mining of semantic and emotional information in social media texts. 2) The design of the neural network model. 3) The combination of emotional information and semantic embedding.

1.2. Motivations and contributions of this paper

The primary purpose of this research is to propose a novel personality detection model based on social media text information, which accurately understands the textual information of the text. We use Bidirectional Encoder Representation From Transformers (BERT) (Devlin, Chang, Lee & Toutanova, 2018) to dynamically encode the text vector. Using the multi-head attention mechanism, each word in the text can be fully taken into account the word vector information of other positions to generate a semantic text encoding. In addition, some studies found that there is a link between personality and emotion (S. Zhao et al., 2019). The model should consider certain text emotional features, which are missing in text encoding. We try to combine the external emotional dictionary to let the text information carry certain emotional information, so as to obtain a more accurate personality detection model. To verify the validity of our work, we utilized two publicly available datasets, the MBTI dataset from Twitter tweets and the Big Five dataset from essays, where each tweet or text was labeled with a psychologist's annotated personality type.

This paper proposes a personality detection model that uses multi-text semantics combined with emotional features for multi-label classification. First, compared with the previous word2Vec and GloVe, we use BERT to embed text information. BERT can effectively generate vector representations of text by the fine-tuned BERT model. We especially concatenate multiple text vectors based on sentence-level representation. Sentence-level representations are generated through self-attention mechanisms and weighted summation token embedding. Second, the SenticNet5[1] emotion dictionary (Cambria, Poria, Hazarika & Kwok, 2018) is used to extract the sentiment polarity of the sentence, then map the sentiment polarity to the vector space, and finally combine this with the text semantic vector as the input to the neural network. Third, compared with the traditional single-text based personality detection model, we concatenate the previous connection of multiple text vectors with the sentiment vector and input them into different neural networks for training and learning. Using a public personality dataset for verification, the results show that our personality detection model can effectively improve personality detection accuracy.

In summary, we perform sentence-level extraction of both semantic and emotion features by using respectively a BERT model and SentiNet5 dictionary. We then apply these features to various neural network architectures (GRU, LSTM, CNN) for further feature extraction and classification. The main contributions of this article are as follows:

1) We propose a novel multi-label personality detection model that combines the pre-trained BERT model and a neural network, which can better understand sentence semantics and process information from social media text data.
2) We propose a method combining semantic and emotional features for the personality detection method, which adds some ability to explain personality and contributes to the analysis of personality traits.
3) Experiment results on the MBTI and Big Five datasets demonstrate that our model outperforms the state-of-the-art techniques in personality detection. Our model can automatically perform personality detection from social media texts, helping various social software for user information mining. Besides, our model can also be applied to other personality classification (e.g., Cattell's 16 Personality Factor) tasks with some simple modifications.

[1] http://www.sentic.net

The rest of the paper is organized as follows. In the next section, Section 2 introduces the related work on personality detection. Section 3 proposes our personality detection method. In Section 4, we introduce two datasets for personality detection and conduct experiments and evaluate the performance of the model. Section 5 discusses this research and the prospects for future research. Finally, Section 6 summarizes this research.

2. Related work

The related work of this paper includes two aspects. The first is personality detection using psychological lexicons, and the second is personality detection using neural networks.

2.1. Personality detection using psychological lexicons

The original method uses a personality trait detection model constructed using the linguistic features of keywords. In the past decade, numerous studies have linked language use with a wide range of psychological correlates. For instance, Park et al. (Park et al., 2015) built a personality prediction model based on the user's language, which indicated that language-based assessment could be a valid measure of personality. Early researchers Pennebaker and King found a reliable association between writing style (e.g., word frequency and part of speech) and personality (Pennebaker & King, 1999), and then they proposed a linguistic inquiry and word count (LIWC) (Pennebaker et al., 2001) method. It first categorizes the words into various psychologically relevant categories, counts the frequency of words in each category, and then predicts the personality traits of the text with the help of some machine learning methods. Afterwards, LIWC was used to construct a web application and Application Programming Interface (API) for personality detection to conduct commercial activities (Golbeck, 2016; Sewwandi et al., 2017). The techniques they proposed can quickly detect a person's personality with high accuracy. Due to the rapid development of the Internet, personality detection from social media texts has become the latest trend. For example, personality detection from the semantic and status updates of social media (Akhtar et al., 2018; Panicheva, Ledovaya & Bogolyubova, 2016), which can help companies select the right employees (Gaddis & Foster, 2015). Similarly, Yin et al. studied the relationship between negative textual information and Big Five personality traits (C. Yin, Zhang & Liu, 2020). They found that user interaction with Weibo for certain personality traits helps predict the spread of negative Weibo messages. All the studies show above show that there is a link between textual features and personality traits. However, due to the dynamic abstract nature of the language used by social media platforms, the interpretation of the results also requires a good psychological language background. Therefore, it is challenging to construct an automatic text-based personality detection model.

In 2013, Poria et al. found that the use of common sense knowledge with affective and sentiment information improves the accuracy of the existing frameworks (Poria, Gelbukh, Agarwal, Cambria & Howard, 2013). They used SenticNet (Cambria et al., 2018), a popular tool used for extracting common sense knowledge along with the associated sentiment polarity and affective labels from the text. These traits are input into a classifier for personality detection. Meanwhile, SenticNet provides the emotional value of the four emotion dimensions (pleasantness, attention, sensitivity, and aptitude) and the emotion polarity values from -1 to $+1$ (where -1 is extremely negative and $+1$ is extremely positive). Darliansyah et al. constructed a sentiment-based personality detection system. Their results indicated that people with the same personality traits tended to express emotions in the same way, which proved the correlation between personality and emotion (Darliansyah, Naeem, Mirza & Pears, 2019). Emotion, as a person's state at a time, is closely related to the long-term expression of personality traits. Therefore, the emotional features extracted by SenticNet can be effectively used for the analysis of personality detection.

2.2. Personality detection using a neural network

The latest method mainly uses deep neural networks for semantic understanding to analyze personality traits. Some review articles give a detailed summary of the field (Mehta, Majumder, Gelbukh & Cambria, 2019; Remaida, Abdellaoui, Moumen & Idrissi, 2020).

2.2.1. Deep learning

After 2014, end-to-end deep neural network architectures and models were developed to extract features more cost-effective methods. Liu et al. (Liu, Preotiuc-Pietro, Samani, Moghaddam & Ungar, 2016) used a word vector generation model based on Bi-RNN to generate word vector representations that were input to a feedforward neural network and used to predict Big Five personalities. Later, Majumder et al. (Majumder, Poria, Gelbukh & Cambria, 2017) used word embedding techniques and convolutional neural networks to automatically dig deeper into the textual semantics from 2467 essays. Rahman et al. compared the effects of different activation functions in convolutional neural networks in personality detection. They found that the overall performance of tanh was better than sigmoid and leaky ReLU when performing personality detection from the text (Rahman, Al Faisal, Khanam, Amjad & Siddik, 2019). The personality traits were detected through a CNN, but this method did not consider the continuity of text information. Hernandez (Hernandez & Knight, 2017) and Scott tried to use Long Short-term Memory (LSTM) (Hochreiter & Schmidhuber, 1997). LSTM was found to give better results compared to a vanilla RNN, GRU, and bi-LSTM. Sun et al. (Sun, Liu, Cao, Luo & Shen, 2018; J.-h. Zhao, Zeng, Xiao, Che & Wang, 2020) introduced abstract feature combinations based on closely connected sentences (Celli, 2012), which they called latent sentence groups, and this combination was effective in improving accuracy (Drexel, 2019). The authors then used bidirectional LSTMs concatenated with a CNN to detect a user's personality traits using the structures of text.

2.2.2. *Word embeddings and transformers*

Word embeddings are the technology of converting some words into N by D-dimensional matrixes, where N is the number of words and D is the dimensional output. In early research, Word2Vec was used for word embedding. In the Word2Vec algorithm, first, a feedforward neural network is trained to predict the next word given a specific word and its previous context. Then, the hidden layer within the feedforward neural network is used to convert the words into related vectors, each word is represented as a fixed-length feature vector with Word2Vec, and the sentence is represented as a variable number of word vectors. Majumder et al. (Majumder et al., 2017) used Word2Vec to embed words, then applied a deep CNN for personality detection from documents. Han et al. (Han, Huang & Tang, 2020) also utilized Word2Vec for textual encoding, and then proposed an interpretable personality model. The Word2Vec algorithm was also applied to train multiple large-scale word embeddings, focusing on large datasets such as Twitter and Wikipedia, and produced pre-trained embeddings like GloVe, a set of word vectors that are trained by over 5 billion articles on online data sources. The GloVe is also widely used for word representation. For instance, Kumar et al. used the GloVe word embedding technique to embed words, and then proposed a personality traits classification system that incorporates language-based features (Kumar & Gavrilova, 2019).

Recent research proposes a new approach to word embedding: the transformer. Instead of using a recurrent neural network, it uses a large number of multi-head attention mechanisms (a method of collecting word context) and feedforward neural networks to improve the current sequence-to-sequence task. In 2018, Google proposed a specific type of transformer: the pre-trained BERT model (Devlin et al., 2018). BERT learns by applying some novel mechanisms (e.g., Masked LM, Next Sentence Prediction) to bidirectionally encoder representation from Transformers to eventually generate more complex word characteristics to compare word semantics. This method can transform knowledge into a vector well. The BERT model is a language representation model that requires super large data, huge models, and huge computational overhead for training. In 2019, Keh et al. (Keh & Cheng, 2019) built a personality detection model using the BERT model for word embedding and they found that textual personality detection based on the BERT model can effectively improve accuracy. Since the context embedding learned by BERT has theoretical and empirical advantages over traditional word embedding, we use the BERT model to construct a new sentence-level embedding in the personality detection model.

3. Proposed method

In this section, we present a multi-label personality detection model based on the pre-trained BERT model and a neural network that combines the semantic and emotional features in the text. As shown in Fig. 1., the model is composed of three following parts:

(1) Sentence embedding: It employs the BERT model for the sentence-level embedding.
(2) Sentiment analysis: We used the SenitcNet5 dictionary for feature extraction.
(3) Neural network classification: The neural network methods used for classification include the CNN and RNN.

3.1. *Sentence embedding*

In order to learn more personality-related features, we leverage the personality dataset to train the fine-tune the BERT model adapted to the domain. Then, we use a fine-tuned BERT model to encode each word in the sample into a vector representation. After the token-level coding representation, each sentence is represented as an (N,768) vector representation, where N is the number of tokens in the sentence. In 2019, Keh et al. used the BERT model for single sentence word embedding to train the personality detection model, but this approach ignored the connection between consecutive social media texts. Inspired by this observation, in order to correlate consecutive data texts, we utilize an approach based on the BERT model for sentence-level embedding coding.

First, we reconsider the personality dataset. We find that each sample data had multiple Twitter tweets or sentences, and most of the current studies use a single Twitter tweet or sentence for personality detection. Personality is a long-term stable characteristic, so it would be a better choice to use the text from multiple consecutive Twitter tweets or sentences of the essays for personality detection.

Each user in the dataset has multiple consecutive Twitter tweets or sentences. Therefore, we plan to use K (the specific values will

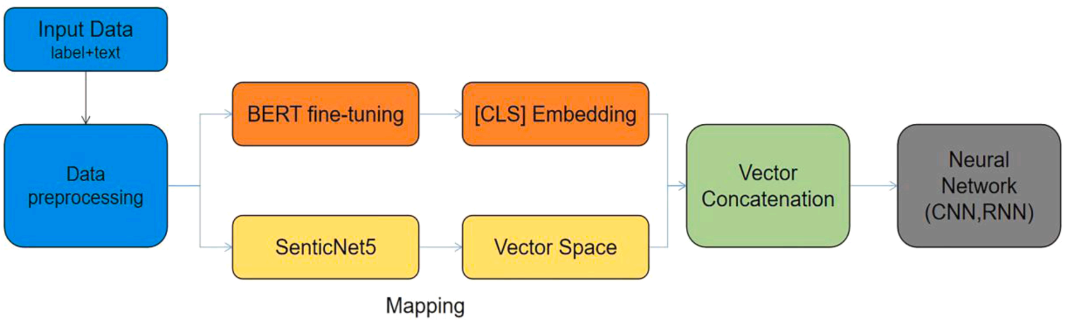

Fig. 1. Experimental flowchart.

be verified in the subsequent experiments) tweets or sentences for personality detection. However, the pre-trained BERT model can accept the text content of up to 512 token sequences, thus the text from K consecutive tweets is too large for text combination by using this model. To solve this problem, we extract a vector named CLS from the penultimate hidden layer of the BERT model to represent each twitter text. We concatenate the CLS vectors of multiple texts instead of the token embedding which takes too much space. The BERT model adds a CLS vector $X \in R^{768}$ at the beginning of each sentence during training. The CLS is weighted through summing up all the token embeddings using self-attention. Thus, we utilize the CLS vector as a semantic representation that can represent a tweet text.

Second, we extract the sentence sequence level representation. Suppose that a sample has K sentences. If there are more than K sentences, the top K sentences are taken. If there are less than K sentences, the 768-dimensional vector is padded with zeros. The final result is: $X \in R^{M*K*768}$, where M represents the number of samples, K represents the maximum length of the sentence, and 768 represents the vector dimension. The main structure of the BERT model for sequence classification tasks consists of 12 hidden layers with a hidden layer size of 768. We can simply understand that each word in the sentence is converted into a 768-dimensional vector representation. So far, we have converted the dataset to a vector representation, and the next step is to train the model. Figs. 2 and 3.

3.2. Sentiment analysis

To further explore and leverage information, sentiment knowledge of the text is a suitable option, and since the current personality dataset does not have any sentiment labeling, it is difficult to use the BERT model for affective feature extraction. The traditional method of emotion feature extraction with the help of external dictionaries is an effective method. LIWC has high confidence in psychology, but because it is not open access, this paper uses an alternative emotion dictionary, SenticNet5, a popular tool used for extracting common sense knowledge along with associated sentiment polarity and affective labels from the text. Thus, we are prepared to train the personality model by combining the emotional and semantic features of the text. Specifically, we use the SenticNet5 sentiment dictionary to detect text words. The SenticNet5 Dictionary can effectively extract the emotional tendency (positive/negative) of certain words and give the corresponding polarity score $x \in (-1, 1)$. If a word is in the sentiment dictionary, we perform sentiment analysis on it. We counted the positive and negative polarity values in a sentence separately and took the average value to represent the emotional polarity of the sentence. First, we append the extracted emotion score as a one-dimensional feature vector to the previous 768-dimensional semantic vector. Through experiments, we found that the results of adding one-dimensional emotional features and not adding emotional features are basically the same. Compared with 768-dimensional semantic features, the emotional features of one-dimensional emotions are too few for neural networks. We decide to map the emotional features to more dimensions. So we try to discretize the emotional polarity score, the scores were divided into twenty rating dimensions from negative to positive (a new sentiment level for every 0.1 increase in the score), and then the emotion scores of these 20 levels were mapped to 20-dimensional feature vectors $x \in R^{20}$. We map each affective rank into a 20-dimensional vector with a value of 1 only at its corresponding position and 0 at all other positions, and then use the results as an indication of the emotional polarity of this sentence, the specific way as Eqs. (1) and Eqs. (2).

$$E_{class} = \left\lfloor \frac{E_{score} + 1}{0.1} \right\rfloor \quad E_{score} \in (-1, 1) \tag{1}$$

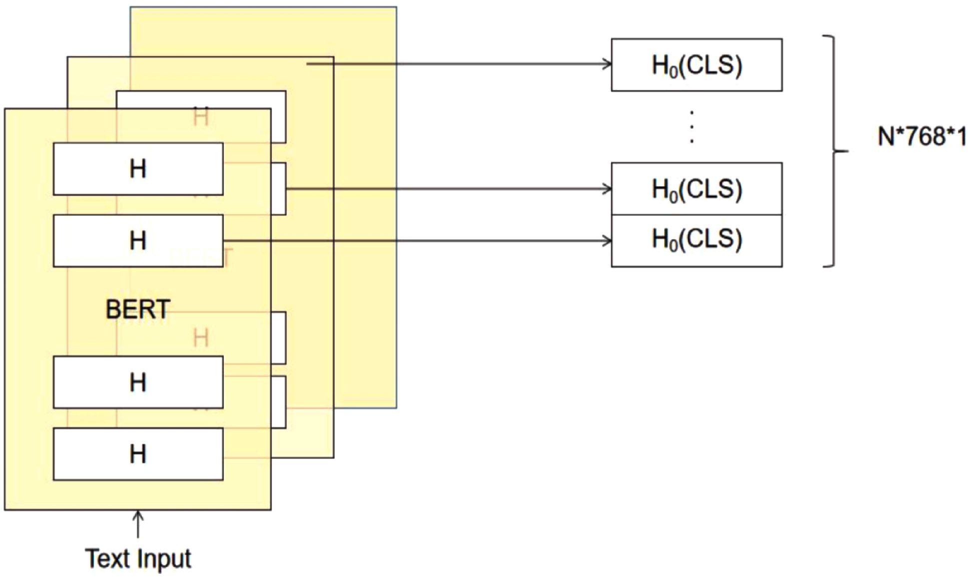

Fig. 2. Extracting textual semantic CLS vector using the BERT.

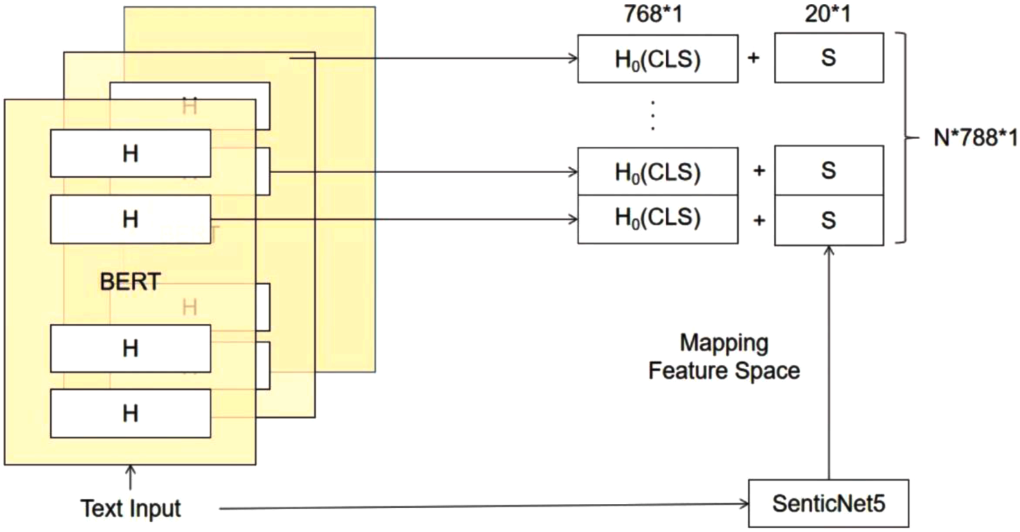

Fig. 3. Building input vectors with SenticNet5.

$$X_{E_{class}} = \left[x^1, x^2, \ldots x^{E_{class}}, \ldots, x^{20}\right] \; x^{E_{class}} = 1, else\ x^i = 0 \tag{2}$$

where E_{score} is the score for each sentiment, E_{class} is the calculated sentiment level, and X_{Eclass} is the final sentiment feature representation.

Afterwards, the emotional polarity vector of each sentence is appended behind the semantic feature vector space and input into the neural network. In the next part, we will introduce our classification neural network.

3.3. Neural network classification

The detection of personality traits can be regarded as a text classification task. At present, the most effective method is deep learning technology. We used three different neural networks in the classification task for the experiments: the CNN, the GRU, and the LSTM. We converted the text samples into sentence embedding and then fed them into the neural network for training.

CNN is a deep learning algorithm that uses the local region view of the previous layer to generate a transmitted intermediate representation for network training. By applying relevant filters, the corresponding input features can be better captured, involving a reduction in the number of parameters and the reusability of weights. In the classification task of text, the sequence of contexts is important, so we also try to use LSTM and GRU. Long short-term memory (LSTM) is a kind of time recurrent neural network, which can bridge minimum time delays of more than 1000 discrete time steps by specifically enforcing a constant error stream (Hochreiter & Schmidhuber, 1997), thus reflecting a dynamic behavior of time sequence. Unlike ordinary RNN, LSTM is specialized at manipulating Long-Term dependencies because it employs the "remember" mechanism through a series of gates while reducing the gradient disappearance problem. GRU is a variant based on LSTM, which aims to speed up the training process while preserving long sequence information.

We resume here the three types of neural networks that will be used:

1) CNN is a convolutional neural network that can effectively perform feature extraction. We used three consecutive convolutional layers for feature extraction. The convolutional kernel sizes are 1×3, 3×3, and 3×1, and then the fully connected layer. The number of output units is equal to the number of personality dimensions, for example, four output units for the MBTI personality dataset and five for the Big Five personality dataset.
2) The GRU is essentially a type of RNN. To explore whether there are some connections between consecutive sentences, we use the neural network structure of the GRU, which can capture the relationships between successive consecutive statements well. We used a two-layer GRU stacked network structure, where the size of all the hidden layer nodes is 100, followed by two successive fully connected layers for training, with the same number of output units as above.
3) LSTM is also used for validation in order to better understand whether the text sequence affects the classification effect. We finally experimented with an LSTM structure with 128 hidden layer nodes as the LSTM neural network, and used the output of the last node to connect the full connected layer for the multi-label classification. The specific network architecture is shown in Fig. 4.

We directly adopt a multi-label classification for model construction and set the number of output units to the number of personality dimensions. All three networks use multi-label classifiers and binary cross-entropy.

$$y = labels \tag{3}$$

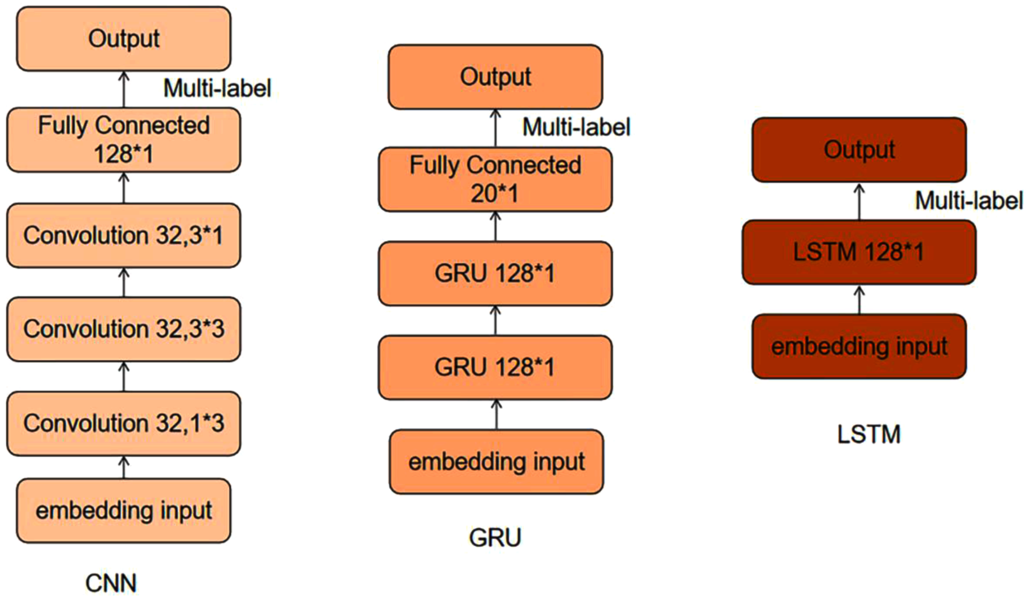

Fig. 4. Neural network model structure.

$$p_{ij} = sigmoid(logits_{ij}) = \frac{1}{1 + e^{-logits_{ij}}} \tag{4}$$

$$loss_{ij} = -[y_{ij} * \ln p_{ij} + (1 - y_{ij})\ln(1 - p_{ij})] \tag{5}$$

In the final network prediction stage, we use the sigmoid function to convert the output scalar number to a range from $[0,1]$. If it is greater or less than a probability threshold (typically 0.5), it is considered to belong to a category.

4. Experiments

In this section, we present the details of the experiment. Two different textual personality datasets are used for the experiments, and the final results are given by comparing them with the state-of-the-art models.

4.1. Personality dataset

Available personality datasets with standard personality labels are rare, and their collection is difficult due to the privacy issues involved, besides the cost of finding professional psychologists for labeling is expensive. Akrami et al. proved that models based on a small high-reliability dataset perform better than models based on a large low-reliability dataset (Akrami, Fernquist, Isbister, Kaati & Pelzer, 2019), so we decide to use some public personality datasets that are more commonly and currently used.

We use two datasets, the MBTI dataset[2] and the Big Five dataset, in the experiment to check the performances of our personality detection model. For the MBTI dataset, some of the recent researches are still using the data (Hernandez & Knight, 2017). Unfortunately, MyPersonality, the largest dataset in the Big Five personality field, was made unavailable in 2018 due to privacy leaks. So we take the dataset from the research of Majumder et al. (Kumar & Gavrilova, 2019; Majumder et al., 2017). The basic dataset information is shown in Table 1.

We resume here the two types of datasets that will be used:

(1) MBTI dataset. The MBTI personality type dataset is a standard dataset and it is the largest publicly published personality dataset that is available. Twitter's MBTI personality dataset, which is divided into 4 different dimensions, is collected through the Personality Café forum. The dataset includes 50 tweet texts from 8675 volunteers and their personality labels. This gives us 422,845 total labeled points in the form (post text, MBTI type). The dataset is available on the Kaggle website[3] and permission has been given for its use in academic research.

[2] https://www.kaggle.com/datasnaek/mbti-type
[3] https://www.kaggle.com

Table 1
Dataset Introduction.

Dataset	Source	Size of dataset	Personality dimensions	Description
MBTI	Twitter	8675×50 (post text, type)	four-dimensional: I-E, S-N, T-F, and J-P	Each item consists of 50 consecutive Twitter texts from the user
Big Five	essays	2467×50 (essay text, type)	five-dimensional: EXT, NEU, AGR, CON, and OPN	Each item includes multiple user sentences

(2) Big Five dataset. The Big Five dataset which we use is James Pennebaker and Laura King's stream-of-consciousness essay dataset (Pennebaker & King, 1999). It contains the content of 2468 anonymous articles with author-tagged Big Five personality dimensions: EXT, NEU, AGR, CON, and OPN. Since one article in the experimental dataset reads "Err:508″, the experiments were conducted using the remaining 2467 pieces of data.

4.2. Data preprocessing

First, we deleted the URL links which begin with the word "http://" or "https://" in the data text, then used the NLTK[4] package to delete some stop words. Finally, we deleted some special symbols that were not letters, numbers, or punctuations. Then, we performed some preliminary statistics on the dataset, the results are as shown in Figs. 5 and 6. In the MBTI dataset, it is obvious that there is a serious data imbalance between EI and SN dimensions. Therefore, we balanced the samples of each dimension in the dataset by undersampling. On the contrary, the classes in the Big Five dataset are more balanced than in the MBTI dataset.

After data preprocessing, our experiment mainly generates two available forms of data for training. The first form of data is to fine-tune the Bert model for domain adaptation. Because the presence of multiple strips of text data per data sample makes it difficult to train the BERT model, we separate each text of a sample as a separate data unit (every tweet or sentence ending with a period, question mark, or exclamation point in essays as a new sample), and the data labels remain unchanged. The second form of data is to treat a person with K consecutive tweets or texts (which have the same label) as a sample, and this is used for all experiments except for fine-tuning the Bert model.

4.3. Experimental setup

4.3.1. Bert fine-tuning

Our experiment is based on the BERT model. The pre-trained BERT model used is the BERT-BASE[5] model (the number of network layers $L = 12$, the hidden layer dimension $H = 768$, attention=12, and the total number of parameters exceed 110 M). Here, we just use our model fine-tuned with our personality data.

We used the first form of data mentioned above (each tweet or sentence as an example) to fine-tune the BERT model Then, we divide the dataset into a training set, validation set, and test set at a ratio of 6:2:2.

After that, we input the data into the BERT model, which is based on the TensorFlow[6] 1.14 framework, with the training parameters as shown in Table 2.

4.3.2. Tuning the consecutive text

In this section, we need to determine the number of consecutive texts, which is the K value mentioned in Section 3.1. In order to verify the specific K, we carried out experiments in two datasets. In the experiment, K starts at 10 and increases by 10 each time, with a maximum value of 50 (the maximum number of consecutive samples of a person in the dataset is 50). K consecutive sentences are encoded by the Bert model. Then they are input to a simple neural network (full connection layer + classification output) for multi-label classification, while a 3 × 3 convolutional kernel is added to the neural network to accelerate the feature extraction. Thus, a BERT+CNN model was used to perform experiments on tuning continuous text. The results are shown in Fig. 7.

In the MBTI dataset, the highest accuracy was achieved between $K = 20$ and 30. After that, we conducted an experiment on $K = 25$, and the results are shown in Table 3. Therefore, we took $K = 25$ on this dataset.

In the Big Five dataset, it can be seen from the figure that with the increase of K value, the accuracy was constantly improved. Therefore, we took $K = 50$ on this dataset.

4.3.3. Pearson correlation coefficient

To investigate the correlation of personality dimensions, we calculated the Pearson correlation coefficients of the personality dimensions of the experimental data. The Pearson correlation coefficient is a statistical method used to measure the degree of correlation between two variables. When the Pearson correlation coefficient is from −0.3 to 0.3(Mukaka, 2012), we can generally consider it to be uncorrelated. In Tables 4 and 5, we can clearly see that there is no correlation between personality dimensions (a maximum of 0.16 and −0.16 between personality dimensions).

[4] http://www.nltk.org
[5] https://storage.googleapis.com/bert_models/2020_02_20/uncased_L-12_H-768_A-12.zip
[6] https://tensorflow.google.cn/

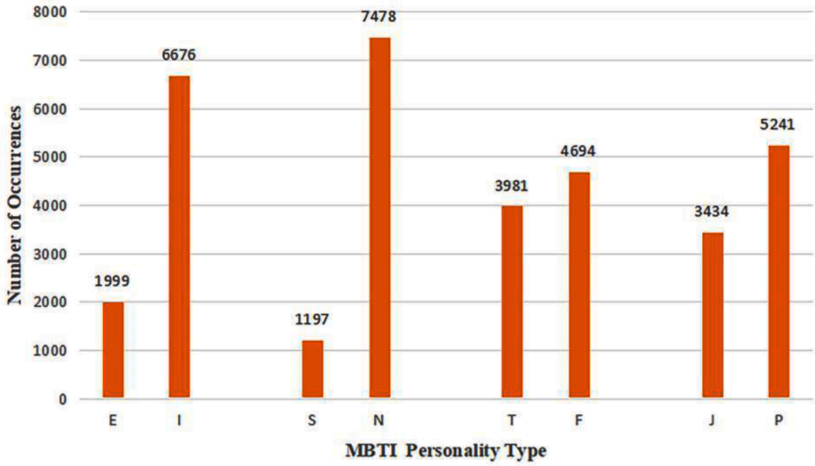

Fig. 5. Statistical charts for MBTI personality types.

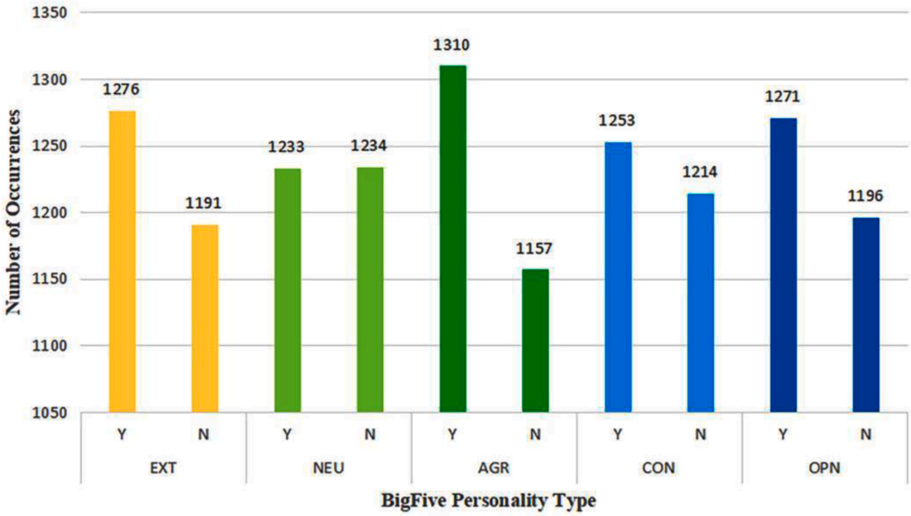

Fig. 6. Statistical charts for Big Five personality types.

Table 2
training parameters of the fine-tuning Bert model.

Dataset	Learning rate	Batch size	Training epochs	Training steps
MBTI	0.00002	16	10	93,404
Big Five	0.00002	16	10	26,557

4.3.4. Training

In the field of personality detection, cross-validation is commonly used in experiments. Cross-fold validation ($k = 4$) is employed to measure the performance of our classifier model. The training parameters are shown in Table 6.

4.4. Comparison with state-of-the-art techniques

We use the following two different datasets for the experiment, and we compared the results from the different datasets separately.

1) MBTI dataset: We compare our approach with the following state-of-the-art techniques. (a) We use the model proposed by Keh et al. in 2019, which is the only current experiment using the BERT model for personality detection. (b) Regarding recent classification techniques based on linguistic features, in 2019, Pavan Kumar K. N et al. proposed a personality detection system that integrates count vectorization (e.g., by TF-IDF) and GloVe word embedding, and this is being used.

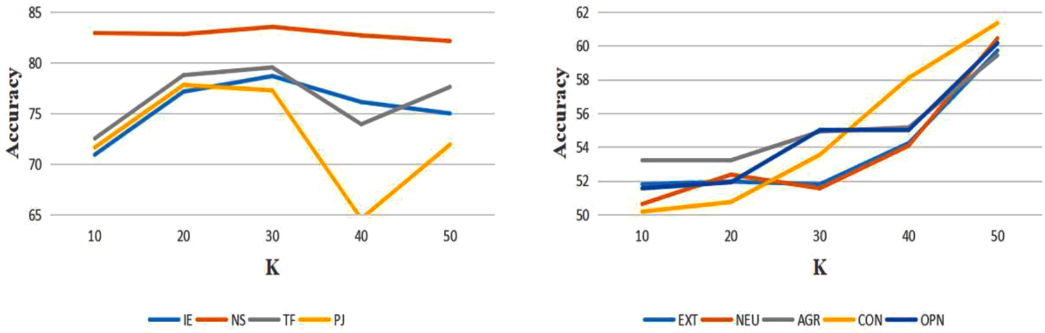

Fig. 7. Different K in MBTI dataset and Big Five dataset.

Table 3
Different K in MBTI dataset.

	IE	NS	TF	PJ
$K = 20$	77.13	82.81	78.76	77.79
$K = 25$	78.84	83.90	78.50	77.88
$K = 30$	78.67	83.52	79.52	77.26

Table 4
MBTI-Pearson coefficients.

	IE	NS	TF	JP
IE	1	−0.046	−0.070	0.160
NS		1	−0.081	0.015
TF			1	−0.00447
JP				1

Table 5
Big Five-Pearson coefficients.

	EXT	NEU	AGR	CON	OPN
EXT	1	−0.16	0.12	0.13	0.079
NEU		1	−0.089	−0.148	−0.047
AGR			1	0.134	0.018
CON				1	−0.027
OPN					1

Table 6
training parameters of classification neural network.

Neural Network	Learning rate	Batch size	Training epochs
CNN/GRU/LSTM	0.001	32	20

Table 7
Single label and multi-label results in MBTI of the BERT+CNN structure.

Personality trait	E-I	S-N	T-F	J-P
BERT (Single label) ($K = 1$)	0.7262	0.7433	0.7502	0.7039
BERT (Multi-label) ($K = 1$)	0.7379	0.7360	0.7706	0.7172
BERT (Single label) ($K = 25$)	0.7842	0.7854	**0.7750**	0.7135
BERT (Multi-label) ($K = 25$)	**0.7884**	**0.8390**	0.7217	**0.7203**

2) Big Five dataset: Since the largest recent MyPersonality dataset is unavailable due to privacy breaches, some research findings based on this dataset cannot be used as a basis for comparison. Therefore, we use the following model. (a) The personality classification technique for deep convolutional neural networks proposed by Majumder et al. in 2017 is being used.

To assess the final classification results, we used the accuracy rate. The accuracy rate is most commonly used in personality trait detection for each personality dimension, and we also used it the same way as the final result for correlation comparisons.

As can be seen from Tables 7 and 8, compared to using the BERT model with single-label for classification detection, the BERT with multiple labels has little or no effect on the results. However, when the K value is different, we can see that the accuracy has improved to a certain extent, so we can infer that continuous text is helpful to the analysis of personality traits. Moreover, in contrast to traditional multi-binary classification models, we only need to train one model of personality detection using multi-label classification.

Tables 9 and 10 show our final results. Compared to traditional neural networks, our model uses the BERT model from the beginning to obtain the embedding vector of the text, thereby taking full account of the contextual content of the text and achieving better results compared to traditional encoding. We then tried to add three different neural networks after the BERT model for deeper trait extraction, and the results are shown in Tables 9 and 10, BERT+CNN obtained the best results and better than using BERT alone. The BERT with a neural network can effectively improve the accuracy of personality detection. Among the three neural networks, CNN gets the best results.

Finally, on the basis of BERT+CNN, we achieved better results by using the embedding vectors of the BERT model combined with SenticNet5 sentiment information. Combining external domain knowledge with text semantics can improve the understanding of the text content and improve accuracy.

The table shows how our results compare to various performance metrics reported in other leading works in the field. In comparison with the BERT model, our model obtained an average accuracy improvement of 6.91% and 6.04% on the two datasets, respectively. Our research shows that the performance of the personality trait detection model based on semantic representation and lexical statistics is superior to traditional language feature-based and neural network methods.

5. Discussion and future work

5.1. Results analysis

Through the analysis of the results, it shows that our personality detection model works better than the state-of-the-art model. Compared to traditional neural networks, our model uses the BERT model to obtain the embedding vector of the text. Traditional coding pales in comparison to the rich BERT word embeddings that are trained utilizing transformers, a coalition of encoders and attention mechanisms. In addition, in order to improve the efficiency of the model, we changed word embedding to sentence embedding, so as to obtain better textual representation. Meanwhile, we changed the classifier to a multi-label classifier. Compared with the traditional multiple binary classification models, we can use the multi-label classification model to obtain personality detection results in multiple dimensions at the same time. Finally, we combined external domain knowledge with text semantics, and by combining the embedding vector of the BERT model and SenticNet5 emotional information, we carried more personality-related information in the input vector of the model to obtain better results.

In the continuous text experiment, we found that the semantic connection of the text helps to obtain the personality results. In the analysis of the specific “K”, we found that in the Twitter text, the first 25 tweets will increase the accuracy, while adding more text connections afterward will decrease the accuracy, which may be related to the periodicity of the text. In the Big Five dataset, the accuracy is improving as the K increases. Among the essays dataset, continuous semantics expresses a clearer meaning, so the more semantic connections lead to better results. From this, we can conclude that semantic continuity helps to increase the accuracy of classification results.

Among CNNs and RNNs, which architecture performs better for text classification tasks depends on how significant the semantic understanding of the whole sequence, and for a classification task such as sentiment classification, it makes sense to choose CNN because sentiment is usually determined by a number of key phrases (W. Yin, Kann, Yu & Schütze, 2017). In the combination of the Bert model and different neural network architectures, CNN performs best. In continuous semantic connections, the overall semantic information is more important. CNN algorithms are good at extracting local and location invariant features, while RNNs are usually good at predicting what will happen next in a sequence, some key semantic information is more important for personality detection than time sequences. CNN can identify the main features of some regions, which makes it can use some key information to classify tasks. Moreover, since we combine different features (semantic, emotional), the input vectors carry rich information. GRU and LSTM can learn information related to time sequences, but they are not as good as CNN in the extraction of some key features. CNN can extract core features of the text efficiently, which leads to better results.

Another interesting result is that MBTI is overall better than Big Five, which is also confirmed by Celli et al. (Celli & Lepri, 2018). Among the MBTI results, the two dimensions of EI and SN are more reliable. Among the Big Five personality data results, there is little difference in the results for each dimension. The results show that our model performs better in the MBTI dataset than the Big Five.

5.2. Implications for research

Our model achieves significant results in personality detection. We gained the following significant implications of the current research: (1) In texts with greater overall semantic relevance, more continuous textual connections can result in a promising outcome. (2) Word embedding and emotional information integration contribute to the effectiveness of personality detection. (3) Emotional

Table 8
Single label and multi-label results in Big Five of the BERT+CNN structure.

Personality trait	EXT	NEU	AGR	CON	OPN
BERT (Single label) ($K = 1$)	0.6290	0.7382	0.6529	0.6929	0.7131
BERT (Multi-label) ($K = 1$)	0.6943	0.6823	0.6971	0.7044	0.6604
BERT (Single label) ($K = 50$)	**0.7180**	0.7016	**0.7268**	**0.7713**	0.7016
BERT (Multi-label) ($K = 50$)	0.7110	**0.7431**	0.7209	0.7073	**0.7485**

Table 9
MBTI dataset results ($K = 25$).

Personality trait	E-I	S-N	T-F	J-P
BERT (2019)	0.7583	0.7441	0.7575	0.7190
TF-IDF+GLOVE+SVM (2019)	0.6830	0.8500	0.7730	0.7560
BERT (Single label)	0.7842	0.7854	0.7750	0.7135
BERT (Multi-label)	0.7884	0.8390	0.7217	0.7203
BERT+GRU (Multi-label)	0.7842	0.8853	0.5839	0.6163
BERT+LSTM (Multi-label)	0.7842	0.8853	0.6716	0.6328
BERT+CNN (Multi-label)	**0.8175**	0.9075	0.7931	0.7876
BERT+Sentic+CNN (Multi-label)	0.8146	**0.9251**	**0.8357**	**0.8236**

Table 10
Big Five dataset results ($K = 50$).

Personality trait	EXT	NEU	AGR	CON	OPN
CNN(2017)	0.5809	0.5938	0.5671	0.5730	0.6268
BERT (Single label)	0.7180	0.7016	0.7268	0.7713	0.7016
BERT (Multi-label)	0.7110	0.7431	0.7209	0.7073	0.7485
BERT+GRU (Multi-label)	0.5253	0.5258	0.5406	0.4983	0.4948
BERT+LSTM (Multi-label)	0.5290	0.5273	0.5399	0.5391	0.5731
BERT+CNN (Multi-label)	0.7630	0.7601	0.7751	0.7634	0.7682
BERT+Sentic+CNN (Multi-label)	**0.7994**	**0.8014**	**0.8030**	**0.8023**	**0.8035**

information is closely related to personality traits and can be attempted to be used to provide further explanations of personality traits. We use effective external knowledge to supplement the information so that our input text carries more information. This is a new attempt, and the combination of deep learning and emotional or mental vocabulary is a future direction worth exploring. (4) Our research introduces a global model based on BERT sentence representation and sentiment detection, which can capture more contextual elements than ordinary word embedding, which should be the basis for future textual personality detection using deep learning.

The study also has significant implications for practice. Social media such as Facebook, Twitter, and Weibo. have become one of the fastest ways to disseminate information. It is essential to identify and target each user individually. Understanding the user's personality traits can be applied to a variety of downstream tasks, such as the accurate recommendation of information, information seeking, etc.

It has been demonstrated in several kinds of research that personality traits affect information-seeking behavior. We can use our model to conduct personality detection and identify different personality traits, which can then be applied to artificial intelligence systems for personalized search settings. We can also combine personality traits with other information to make recommendations, such as clustering people with the same personality traits and preferences. Therefore, an accurate model of personality detection is necessary.

5.3. Future research

In our experiments, the dataset is relatively small, and the future application of big data analysis is the direction we should consider. Many factors affect personality detection and we are only using textual information to analyze personality. In fact, other auxiliary information such as audio and video would help us make more accurate predictions. It is challenging to design a completely accurate personality detection system. In the future, multimodal convergence will be the dominant technology in this field. We will combine the image information of social media in subsequent research for further personality analysis.

6. Conclusion

In this paper, we propose a novel personality detection model based on emotional and semantic features, which can perform accurate personality detection using just a few social media texts. We first utilize the pre-trained BERT model, a cutting-edge technical approach in the field of NLP, for sentence-level text semantic extraction. Then, we use a sentiment dictionary, which adds some ability to explain personalities, to analyze the sentiment of each text and map it to the vector space. Finally, we combine a user's multiple textual contents for combinatorial analysis. The two vectors, emotional vectors and semantic vectors, are input into a neural network classification model to compute the results using a multi-label classifier. Our model is validated on the MBTI and Big Five datasets and can generate superior personality detection results. Therefore, the use of NLP techniques to extract semantic features combined with traditional psychological lexicons such as sentiment analysis is promising.

Declaration of Competing Interest

The authors certify that there is no conflict of interest in the subject matter discussed in this manuscript.

CRediT authorship contribution statement

Zhancheng Ren: Conceptualization, Methodology, Software, Data curation, Writing - original draft, Writing - review & editing. **Qiang Shen:** Visualization, Investigation. **Xiaolei Diao:** Software, Validation. **Hao Xu:** Supervision.

Acknowledgements

This research is supported by the National Natural Science Foundation of China (62077027), the Ministry of Science and Technology of the People's Republic of China (2018YFC2002500), the Jilin Province Development and Reform Commission, China (2019C053–1), the Education Department of Jilin Province, China (JJKH20200993K), and the Department of Science and Technology of Jilin Province, China (20200801002GH), the European Union's Horizon 2020 FET Proactive project "WeNet-The Internet of us" (No 823783). Thanks to the International Innovation Team of Philosophy and Social Science of Jilin University for some support.

References

Akhtar, R., Winsborough, D., Ort, U., Johnson, A., Chamorro-Premuzic, T. J. P., & Differences, I. (2018). Detecting the dark side of personality using social media status updates. *Personality and Individual Differences, 132*, 90–97.
Akrami, N., Fernquist, J., Isbister, T., Kaati, L., & Pelzer, B. (2019). Automatic extraction of personality from text: Challenges and opportunities. In *Paper presented at the 2019 IEEE International Conference on Big Data (Big Data)*.
Al-Samarraie, H., Eldenfria, A., & Dawoud, H. (2017). The impact of personality traits on users' information-seeking behavior. *Information Processing & Management, 53*(1), 237–247.
Azucar, D., Marengo, D., & Settanni, M. (2018). Predicting the Big 5 personality traits from digital footprints on social media: A meta-analysis. *Personality and Individual Differences, 124*, 150–159.
Boxman-Shabtai, L., & Shifman, L. (2014). Evasive targets: Deciphering polysemy in mediated humor. *Journal of Communication, 64*, 977–998.
Briggs Myers, I., & Kirby, L. K. (2000). *Introduction to type: A guide to understanding your results on the Myers-Briggs type indicator*. European English Version.
Cambria, E., Poria, S., Hazarika, D., & Kwok, K. (2018). Paper presented at the. In *Thirty-Second AAAI Conference on Artificial Intelligence*. Paper presented at the.
Celli, F. (2012). In *Paper presented at the Proc. of sixth international conference on digital society*.
Celli, F., & Lepri, B. (2018). *Is big five better than MBTI? a personality computing challenge using twitter data*. Paper presented at the CLiC-it.
Darliansyah, A., Naeem, M. A., Mirza, F., & Pears, R. (2019). SENTIPEDE: A Smart System for Sentiment-based Personality Detection from Short Texts. *J. UCS, 25*(10), 1323–1352.
Devlin, J., Chang, M.-. W., Lee, K., & Toutanova, K. (2018). *Bert: Pre-training of deep bidirectional transformers for language understanding*.
Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology, 50*(1), 417–440.
Drexel, I. B. (2019). In *Paper presented at the International Conference on Information Management*.
Elman, J. L. (1990). Finding Structure in Time. *Cognitive Ence, 14*(2), 179–211.
Fukushima, K., Miyake, S., & Ito, T. (1988). Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Visual Pattern Recognition. *IEEE Transactions on Systems Man and Cybernetics, SMC-13*(5), 826–834.
Gaddis, B. H., & Foster, J. L. (2015). Meta-analysis of dark side personality characteristics and critical work behaviors among leaders across the globe: Findings and implications for leadership development and executive coaching. *Applied Psychology, 64*(1), 25–54.
Golbeck, J. A. (2016). Predicting personality from social media text. *AIS Transactions on Replication Research, 2*(1), 2.
Han, S., Huang, H., & Tang, Y. (2020). Knowledge of words: An interpretable approach for personality recognition from social media. *Knowledge-Based Systems*, Article 105550.
Hernandez, R., & Knight, I. (2017). In *Paper presented at the Proceedings of the 31st Conference on Neural Information Processing Systems*.
Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780.
John, O. P., Robins, R. W., & Pervin, L. A. (2010). *Handbook of personality: Theory and research*. Guilford Press.
Keh, S.S., .& Cheng, I.T. (.2019). Myers-Briggs personality classification and personality-specific language generation using pre-trained language models. arXiv preprint arXiv.06333.
Kumar, K. N. P., & Gavrilova, M. L. (2019). In *Paper presented at the 2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*.
Liu, L., Preotiuc-Pietro, D., Samani, Z. R., Moghaddam, M. E., & Ungar, L. H. (2016). *Analyzing personality through social media profile picture choice*. Paper presented at the ICWSM.
Majumder, N., Poria, S., Gelbukh, A., & Cambria, E. (2017). Deep learning-based document modeling for personality detection from text. *IEEE Intelligent Systems, 32*(2), 74–79.
Matthews, G., Deary, I., & Whiteman, M. (1998). *Personality trait*. Cambridge, UK: Cambridge University Press.
Mehta, Y., Majumder, N., Gelbukh, A., & Cambria, E. (2019). Recent Trends in Deep Learning Based Personality Detection. *Artificial Intelligence Review*.
Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *Computer ence*.

Mukaka, M. (2012). Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi Medical Journal : The Journal of Medical Association of Malawi, 24*(3), 69–71.

Panicheva, P., Ledovaya, Y., & Bogolyubova, O. (2016). *Lexical, morphological and semantic correlates of the dark triad personality traits in russian facebook texts*. Paper presented at the 2016 IEEE artificial intelligence and natural language conference (AINL.

Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., & Stillwell, D. J. (2015). Automatic personality assessment through social media language. *Journal of Personality Social Psychology, 108*(6), 934.

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

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality Social Psychology, 77*(6), 1296.

Pennington, J., Socher, R., & Manning, C. D. (2014). In *Paper presented at the Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*.

Poria, S., Gelbukh, A., Agarwal, B., Cambria, E., & Howard, N. (2013). In *Paper presented at the Mexican International Conference on Artificial Intelligence*.

Rahman, M. A., Al Faisal, A., Khanam, T., Amjad, M., & Siddik, M. S. (2019). In *Paper presented at the 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*.

Remaida, A., Abdellaoui, B., Moumen, A., & Idrissi, Y. E. B. E. (2020). In *Paper presented at the 2020 1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*.

Russell, J. D., Weems, C. F., Ahmed, I., & Richard, G. G. (2017). Self-reported secure and insecure cyber behaviour: Factor structure and associations with personality factors. *Journal of Cyber Security Technology*, 1–12.

Sewwandi, D., Perera, K., Sandaruwan, S., Lakchani, O., Nugaliyadde, A., & Thelijjagoda, S. (2017). In *Paper presented at the 2017 6th National Conference on Technology and Management (NCTM)*.

Sun, X., Liu, B., Cao, J., Luo, J., & Shen, X. (2018). In *Paper presented at the 2018 IEEE International Conference on Communications (ICC)*.

Taramigkou, M., Apostolou, D., & Mentzas, G. (2018). Leveraging exploratory search with personality traits and interactional context. *Information Processing & Management, 54*(4), 609–629.

Yin, C., Zhang, X., & Liu, L. (2020). Reposting negative information on microblogs: Do personality traits matter? *Information Processing & Management, 57*(1), 1–18.

Yin, W., Kann, K., Yu, M., & Schütze, H. (2017). Comparative study of cnn and rnn for natural language processing. arXiv preprint arXiv:01923.

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

Zhao, S., Gholaminejad, A., Ding, G., Gao, Y., Han, J., & Keutzer, K. (2019). Personalized emotion recognition by personality-aware high-order learning of physiological signals. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 15*(1), 1–18.