



Contextualized Multidimensional Personality Recognition using Combination of Deep Neural Network and Ensemble Learning

Fatemeh Mohades Deilami¹ · Hossein Sadr² · Morteza Tarkhan³

Accepted: 28 February 2022 / Published online: 5 April 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Personality is generally expressed as the combination of behavior, emotion, motivation, and thoughts that aim to describe different aspects of human behavior based on a few stable and measurable characteristics. Considering the fact that our personality has a remarkable impact in our daily life, automatic recognition of a person's personality attributes can provide many essential practical applications in various aspects of cognitive science. Although various methods have been recently proposed for the task of personality recognition, most of them have mainly focused on human-designed statistical features and they did not make use of rich semantic information existing in users' generated texts while not only these contents can specify its writer's internal thought and emotion but also can be assumed as the most direct way for people to state their feeling and opinion in an understandable form. In order to make use of contextualized information as well as overcoming the complexity and handcraft feature requirement of previous methods, a deep learning based method for the task of contextualized personality recognition is proposed in this paper. Among various deep neural networks, Convolutional Neural Networks (CNN) have demonstrated profound efficiency in natural language processing and especially personality detection. Owing to the fact that various filter sizes in CNN may influence its performance, we decided to combine CNN with AdaBoost, a classical ensemble algorithm, to consider the possibility of using the contribution of various filter lengths and gasp their potential in the final classification via combining various classifiers with respective filter size using AdaBoost. Our proposed method was validated on the Essay and YouTube datasets by conducting a series of experiments and the empirical results demonstrated the superiority of our proposed method on both datasets compared to both machine learning and deep learning methods for the task of personality recognition.

Keywords Deep learning · Convolutional neural network · AdaBoost · Cognitive science · Contextualized personality recognition · Ensemble learning

✉ Hossein Sadr
Sadr@qiau.ac.ir

¹ Department of Computer Engineering, Ayandegan Institute of Higher Education, Tonekabon, Iran

² Department of Computer Engineering, Rahbord Shomal Institute of Higher Education, Rasht, Iran

³ Department of Psychology, Payam Noor University, Tehran, Iran

1 Introduction

Personality is a stable tendency and characteristic that determines the similarities and differences in psychological behaviors (thoughts, emotions, and actions) of individuals. In other words, personality not only shows an individual's behavioral pattern, thought, and interpersonal communications but also has a great impact on various life aspects, such as happiness, preference, physical, and mental health [1–3]. The evolution of personality theories and their progress in measurement methods and statistical analysis has led to the emergence of one of the most influential contemporary personality theories, known as the Five Factor Model (FFM) or Big Five, which contains five primary traits and has been considered by many psychologists in recent years as a popular and powerful approach for studying personality traits [4, 5]. According to the FFM, personality consists of five main dimensions including neuroticism (NEU), extraversion (EXT), openness to experience (OPN), agreeableness (AGR), and conscientiousness (CON). It is worth mentioning that the study of personality is not only essential for psychology and personality recognition but also can benefit various applications, such as cognitive science [6], social network analysis [7], sentiment analysis [8–10], recommender systems [11], deception detection [12], and so on. In this regard, personality computing is now known as field which investigates the combination of psychology and machine learning as a computational model for understanding personality. Nevertheless, it must be noted that building a general system for this aim is very challenging due to the uncertain delineation of human behaviors in the diverse situation. However, considering the importance of automatic personality recognition, numerous studies focusing on context-specific problems have been conducted in this field over the past few years [13, 14].

On the other hand, by the development of the online social network, several studies have focused on the text generated by people in social media for predicting personality. Despite other the existing methods that commonly utilized questionnaire investigations or expert reviews that were not only costly, time-consuming, and less practical but also were highly dependent on an expert and only utilized human-designed statistical features to perform recognition and did not consider the valuable information existing in texts, these studies only considered text as the most direct way of expressing thought and emotion while it contains rich self-disclosed personal information that is highly correlated with people's personality and interpreting them can provide us with valuable information about users' behavior and feelings [15, 16].

By taking the significance of textual data into account, a small number of studies have focused on using text generated by people to predict their personalities [17]. In this regard, machine learning based methods have been also utilized but their obtained results were not satisfactory because the majority of them were based on statistical or hand-craft linguistic features and were not able to consider the rich user-generated textual information and extract features from them automatically while these words and text are the most valuable features for determining the emotion and personality [18, 19].

By the development of deep neural networks, they demonstrated remarkable performance in various Natural Language Processing (NLP) tasks including opinion mining and sentiment analysis [8, 9]. It must be noted that personality recognition is very similar to NLP applications while they both focus on mining users' attributes from texts. Accordingly, employing powerful text modeling techniques that have been efficiently utilized in the NLP domain can be the most intuitive and straightforward idea for improving the performance of personality recognition [15, 20]. However, capturing potential and efficient features from the text that have a close relation to one's personality is still challenging.

Having the mentioned limitations besides the potential of deep learning in our mind, we proposed a deep learning based method for personality recognition that tries to make use of both Convolutional Neural Network (CNN) and AdaBoost algorithm [21]. CNN has been successfully utilized for various NLP tasks and extracting local features can be considered as its potential. Moreover, CNN generally utilizes filters of various lengths to generate feature maps. Due to the fact that different sizes of filters can lead to the generation of various kinds of N-grams that can help differently interpret and parse sentences, features obtained from various filters can also play an important role in predicting the personality. To this end, we decided to combine CNN with AdaBoost algorithm to investigate the possibility of leveraging the contribution of different filter lengths and grasp their potential for personality recognition by combining classifiers with respective filter sizes. The reason behind choosing AdaBoost is that it is a Meta algorithm that can be used in conjunction with other learning algorithms to improve classification accuracy. Based on this algorithm, the classification of each new stage is adjusted in favor of incorrectly classified samples in the previous stages. In fact, with the help of AdaBoost algorithm, the classification process is repeated until the classification error is minimized.

To prove the efficiency of our proposed method, it was examined on two heterogeneous datasets, namely Stream-of consciousness essays and YouTube personality dataset, and obtained better results compared to other methods which demonstrate that our proposed method can not only capture more advanced structural features but also is more efficient in detecting users' personality traits. In summary, the contributions of this paper are as follows:

- (1) We designed a new structure based on the integration of CNN and AdaBoost for predicting personality from texts where different lengths and various weigh matrices are used in the convolutional layer to extract features. To the best of our knowledge, it is the first study that tries to investigate the combination of CNN and AdaBoost for the task of personality recognition.
- (2) We considered different variations of the proposed method, derived from various vector representation models, in our implementation to prove the efficiency of vector representation in personality recognition.
- (3) Despite other existing studies, we conducted our experiments on two different datasets to clearly explain the versatility and generalization of our method, and based on the empirical result, the proposed method demonstrated higher efficiency compared to both machine learning and deep learning based methods.

The remainder of this paper is organized as follows: Related studies with a focus on deep learning-based methods are presented in Sect. 2. Section 3 includes the details of the proposed methods. Experimental details and obtained results are extensively reported in Sect. 4. Conclusion and possible future directions are also mentioned in Sect. 5.

2 Related Work

Along with the explosive popularity of social media, various studies have been also conducted for personality recognition. Personality recognition methods can be generally divided into two major categories of psychology-based and artificial intelligent based methods [15, 22]. From the perspective of psychology, personality theorists have developed unique methods for assessing individuals' personalities and by applying these methods; they obtained valuable information and then based their formulations on it. On the other hand, artificial intelligence-based methods combined with psychology and machine learning as a computational model

for understanding personality. Given that the focus of this paper is on automatic personality recognition from texts, especially people's opinions about various topics, the present study falls into the group of artificial intelligent based methods. Artificial intelligent based methods for personality recognition are divided into two groups of machine learning and deep learning based methods. More details about these methods and related studies are reported in the following.

2.1 Machine Learning based Methods

With the rapid development of the Internet and social media, numerous studies have been conducted on personality recognition from text. In this regard, Golbeck et al. [23] used M5' rule and Gaussian process for predicting personality based on Big Five scores. They utilized 167 Facebook users' personal information, activity, preferences, and language usage to extract 77 features and perform classification. Following a similar line of research, Golbeck et al. [24] utilized 297 Twitter users' information and applied a similar approach to predict the personality. Moreover, the relation between personality and various kinds of users was also analyzed by Quercia et al. [25]. They used M5' algorithm to predict 335 Twitter users' personalities based on Big Five attributes according to their number of followers, following, and listed counts. Alam et al. [26] utilized a bag of word methods besides unigrams as features to perform personality recognition. They applied various techniques, such as Support Vector Machine (SVM), Multinomial Naïve Bayes (MNB), and Bayesian Logistic Regression (BLR) to predict Big Five attributes according to My personality corpus. Skowron et al. [27] collected text, image, and users' metadata from Instagram and Facebook and performed various machine learning techniques to predict personality. They concluded that joint analysis could enhance performance. Li et al. [28] also proposed a semi-supervised method that utilized over 547 Chinese active users of SinaWeibo to predict personality.

In the following, Bai et al. [29] utilized the information of 209 users of PenPen (Chinese social network) to predict their personality. They analyzed various attributes including usage statics, emotional state, and demographic information, and then applied C4.5 decision tree to perform classification. Peng et al. [30] utilized SVM to predict the personality of 222 Chinese Facebook users based on their generated texts about various topics. Argamon et al. [31] used word categories and relative frequency of function words as SVM input to make discrimination between students at the opposite extremes of neuroticism and extraversion. The efficiency of various textual features extracted from the psycholinguistic dictionary or psychologically oriented text analysis tools was also explored by Mairesse et al. [32] N-gram frequency was another feature that was commonly used as the input of SVM or Naïve Bayes for classifying low and high scoring blogger for Big Five personality attributes.

As it is clear, a large number of studies have focused on traditional machine learning methods to perform personality recognition and the majority of them were highly dependent on the handcraft features like online activities, profile information, or manually extracted features of texts. In other words, the machine learning based methods required an expert to extract features and were not able to make use of rich features existing in the text.

2.2 Deep Learning based Methods

By the rapid growth of deep learning, deep neural networks obtained remarkable efficiency in various NLP tasks. Due to the fact that personality recognition from text is very similar to other NLP tasks like text classification or sentiment analysis, deep neural networks have

also found their way in personality recognition. In this regard, Jiango et al. [33] utilized deep learning methods to predict the personality of Facebook users. They applied a fully connected network, Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) in their experiments and proved the superiority of deep learning based methods for personality recognition compared to other existing methods. In similar research, Tandra et al. [34] utilized multilayer perceptron, Long Short Term Memory (LSTM), Gated Recurrent Unit (GRU), and 1-DCNN to predict Facebook users' personality according to Big Five personality attributes. Similarly, Majumder et al. [5] employed CNN for extracting deep semantic features and predicting the personality based on them. Xue et al. [4] also proposed AttCNN model to extract deep semantic features from users' post text and then concatenated them with statistical linguistic features. They fed the obtained features to a regression algorithm to predict the personality based on the Big Five personality attributes. It is worth mentioning that although deep neural networks have been rarely employed for the task of personality recognition, they have obtained considerable results and they are actually in the early steps of their development and growth.

3 Proposed Methodology

Due to the fact that language is the most reliable way for people to state their opinion and internal feeling in an understandable way, it can be considered as valuable knowledge for psychologists to interpret people's feelings and predict their personalities. In other words, while text can reflect various aspects of its author, efficiently modeling the text generated by authors can improve the performance of personality recognition [35]. Motivated by this intuition, we decided to use a reinforced convolutional neural network architecture with various filters to perform classification. Unlike previous convolutional neural network-based personality recognition methods that combined the properties of different filters into a unified vector that was then fed to a fully connected network to predict the personality [4, 5], various features obtained from various filters of the convolutional neural network are fed to a separate pooling and classifier in our proposed method. When the initial results are obtained by each of the classifiers, the AdaBoost algorithm [21] is used to produce the overall classification results. The reason for choosing AdaBoost is that it is a Meta algorithm that can be used in conjunction with other learning algorithms to improve performance. In this algorithm, the classification of each new sample is adjusted in favor of incorrectly classified samples of other classifiers. In other words, AdaBoost is able to combine weak classifiers with a strong classifier because it can learn the classification error of each weak classifier and accordingly adjust the weight of the classifier for the final classification.

The proposed method includes five steps and it is a combination of convolutional neural network and Adaboost algorithm [21]. In the proposed method, different filters with various sizes are used to scan the input sentence and extract precious low-level features from the input text. In other words, each convolutional neural network has its own unique convolutional, pooling, and classification layer, and classification is performed separately in each convolutional neural network. Finally, the AdaBoost aggregation algorithm is used to create a robust classification based on the different weights of the different classifiers and the personality type is then estimated based on it.

In general, the idea of this paper is based on this hypothesis that using various filters with convolutional neural networks leads to the generation of different features that each of them may have a different effect on the final classification. In the classical convolutional neural

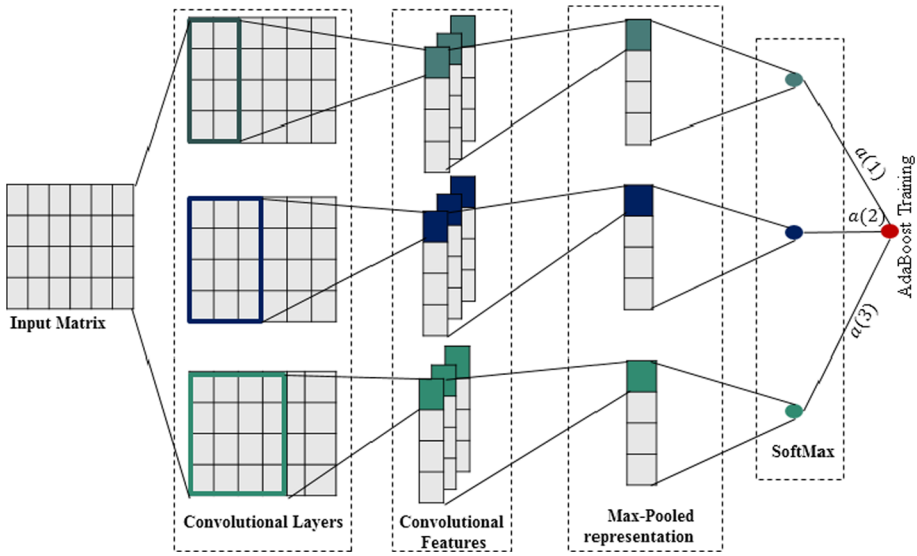


Fig. 1 Integration of convolutional neural network with AdaBoost for personality recognition

networks, these features are merged after applying the pooling operation and classification is performed on the merged features. This can eliminate the impact of some features that may contain valuable information. In this regard, we decided to feed the features obtained from each filter to a separate pooling and classification layer. Next, the classification results are then combined using the AdaBoost algorithm to obtain the best final. The schematic structure of the proposed method is illustrated in Fig. 1.

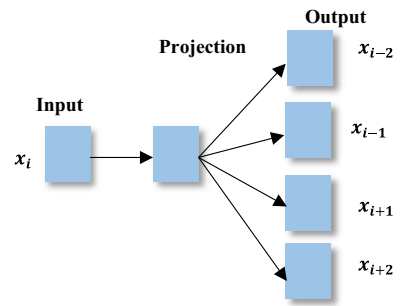
3.1 Representation Layer (Word Matrix Formation)

Data representation refers to language modeling techniques in natural language processing that aims to map words from a very large space to a continuous vector space with much smaller dimensions. In other words, in order to be able to apply a deep learning method for the task of text classification, the words must be transformed into high dimensional vectors to capture the syntactic, semantic, and morphological information of the words. To this end, Skip-Gram [36] model which is a shallow two layers neural network and tries to learn vector representation of a word based on its context is used in the first layer of our proposed method. The diagram of the Skip-Gram model is depicted in Fig. 2 and its object is to find word representations that are efficient for the prediction of the surrounded words in a sentence. Let $x_1, x_2 \dots x_n$ is a sequence of training words, Skip-Gram aims to maximize the average log probability (Eq. 1).

$$\frac{1}{n} \sum_{i=1}^n \sum_{-c \leq j \leq c, j \neq 0} \log p(x_{i+j}|x_i) \tag{1}$$

where c refers to the training context size while a larger value of c yields to more training samples and higher accuracy. Noteworthy, the value of $p(x_{i+j}|x_i)$ is also obtained using SoftMax function. Finally, considering that n is the size of the vocabulary and d is the size

Fig. 2 Skip-Gram model structure



of word embedding, each word is encoded by a column vector in $A \in \mathcal{R}^{n \times d}$ as the sentence matrix.

3.2 Convolutional Layer

The objective of the convolutional layer is to extract local features as well as retaining the sequential information of the input text. To this end, the obtained sentence matrix $A \in \mathcal{R}^{n \times d}$ is fed to the convolutional layer to produce new features. According to the fact that the sequential structure of a sentence has an important effect in specifying its meaning, it is sensible to choose filter width equal to the dimensionality of word vectors (d). In this regard, only the height of filters (h), known as region size, can be varied.

Considering $A \in \mathcal{R}^{n \times d}$ as a sentence matrix, convolution filter $H \in \mathcal{R}^{h \times d}$ is applied on A to produce its submatrix as a new feature $A[i : j]$. As the convolution operation is applied repeatedly on the matrix of A , $O \in \mathcal{R}^{n-h+1 \times d}$ as the output sequence is achieved (Eq. 2).

$$O_i = w \cdot A[i : i + h - 1] \quad (2)$$

here $i = 1, \dots, n - h + 1$ and \cdot is the dot product between two matrices of the convolution filter and input submatrix. Bias term $b \in \mathcal{R}$ and an activation function are also added to each O_i . Finally, feature maps $C \in \mathcal{R}^{n-h+1}$ are generated (Eq. 3).

$$C_i = f(O_i + b) \quad (3)$$

3.3 Pooling Layer

While various feature maps according to different filter sizes are generated, a pooling function is required to induce fixed size vectors. Various strategies such as average pooling, minimum pooling, and maximum pooling can be used for this aim and the idea behind them is to capture the most important feature from each feature map and reduce dimensionality. Maximum pooling is used in our proposed method (Eq. 4).

$$c_{max} = \max\{C\} = \max\{c_1, \dots, c_{n-h+1}\} \quad (4)$$

It is worth mentioning that the pooling layer makes the proposed method aware of the order of the sentences and distributes information related to the personality of the individuals throughout the sentence. On the other hand, the pooling layer allows us to work with sentences of variable length, given that the number of features in the proposed method is aligned with the number of filters. Moreover, the pooling layer reduces the size of the feature maps and

future computations. Features obtained from the pooling layer are then processed using a nonlinear function before being classified.

3.4 Regularization Layer and SoftMax

In order to overcome overfitting, which is known as one of the most important weaknesses of neural networks, dropout is used as a regularization technique in our proposed method. Based on this technique, the values of some features are set to zero. It means that if $C_{max} = \{c_{max}^1, c_{max}^2, \dots, c_{max}^m\}$ are features obtained from the previous layer (m is the number of filters in the convolutional layer), some of them are randomly set to zero before the SoftMax layer. Notably, the dropout value is a hyper-parameter that is specified along with the training. The classification result is the SoftMax output after the regularization layer. This layer employs the regularized features as the input of SoftMax layer to calculate the probability of distribution over all five different kinds of personality based on Big Five attributes (Eq. 5).

$$P(y = j|x) = \text{softmax}_j(x^T w + b) = \frac{e^{x^T w_j + b_j}}{\sum_{k=1}^K e^{x^T w_k + b_k}} \quad (5)$$

where w_k is the input weight, b_k is the bias term, P_i refers to the output class and K is the number of output classes.

3.5 AdaBoost Training Layer and Prediction Integration

AdaBoost is an algorithm that tries to integrate weak classifiers into a strong classifier. Accordingly, we employed this algorithm in our proposed method to find the appropriate weights for the classifiers adjusted to different N-grams. In this regard, there is a need to obtain the statistics of the weak classifier results based on the training samples and then adjust the weights of training samples and classifiers to achieve the final strong classification. Backpropagation is also used to train the network ahead of the AdaBoost integrating part. Training process of AdaBoost can be stated as follows:

1. Initialize equal distribution of \mathcal{D}^1 to all training samples while \mathcal{D}^i specifies the i th training sample distribution (Eq. 6)

$$\mathcal{D}_i^1 = \frac{1}{\#training_samples} \quad (6)$$

2. In each training epoch of t :

While backpropagation is applied to train three neural networks consecutively, the following process is performed on all classifiers in each epoch.

- a. Estimating weak classifiers statistics: After training the classifiers and predicting the output labels, classification statistics over the samples are saved, and weak classifier error $G_m(x)$ is then calculated (Eq. 7).

$$e_m^t = \sum_i \mathcal{D}_i^t 1(G_m(x) \neq y(x)) \quad (7)$$

- b. Adjusting weight: As a weak classifier is trained, the classification error is used to modify the distribution over the training set. Thereafter, the error index and weak classifiers' weights are calculated.

- Calculating classifier weights (Eq. 8):

$$a(m) = \frac{1}{2} \ln \frac{1 - e_m^t}{e_m^t} \quad (8)$$

- Adjusting distribution (Eq. 9):

$$\mathcal{D}_i^{t+1} = \frac{\mathcal{D}_i^t \exp(-a(m)y(x)G_m(x))}{\mathcal{D}^t} \quad (9)$$

- c. Improved validation: As the training process is finished, element-wise multiplication of weights and outputs is performed to obtain the final predicted class of the personality. The learned weight a is then used to perform the improved validation using the following equation (Eq. 10) where i is the classifier index, l refers to the output label of the classifier and a determines the ensembles of the classifier weights.

$$L(s) = \sum_i a(i) * l(i) \quad (10)$$

4 Experiments

Experiments that were carried out to prove the efficiency of the proposed method and obtained results are explained in this section in detail. It is worth mentioning that two datasets were leveraged in our experiments that are extensively introduced in the following.

4.1 Dataset

In order to prove the versatility and generalization of our proposed method, two heterogeneous personality recognition datasets were used in our experiments.

- (1) Stream-of consciousness essays.

Essay is a large dataset based on the stream of consciousness that was collected by James Pennebaker and Laura King according to the text generated by 2467 users between 1997 and 2004 that were labeled based on classes of personality traits including neuroticism (NEU), extraversion (EXT), openness to experience (OPN), agreeableness (AGR), and conscientiousness (CON)). Therefore, the dataset includes a label for each essay indicating the personality of its author and can be suitable for supervised learning. It is worth mentioning that the texts of this dataset were generated by students of the American Psychological Association.

- (2) YouTube personality dataset.

This dataset comes from about 400 YouTube bloggers' webcam videos containing speech transcriptions, gender, and behavioral features translated manually from the videos. In contrary to the first dataset, this dataset contains shorter text, and personality impressions (labels) are determined from the annotator rating impressions by watching each blog.

It is worth mentioning that we decided to examine our proposed method on two datasets to test if our method can handle various cases. The number of statistics of these two datasets are summarized in Table 1. The reason behind choosing these two datasets are as follows:

- As it is clear in Fig. 3, YouTube dataset documents are shorter than Essay dataset which can verify if our proposed method is effective for both long and short texts.

Table 1 Summary statistics of essay and YouTube datasets

Property	Essay	YouTube
Average word count	648	562
Max word count	2488	1972
Min word count	33	36
Average sentence count	46	41
Max sentence count	327	147
Min sentence count	1	2

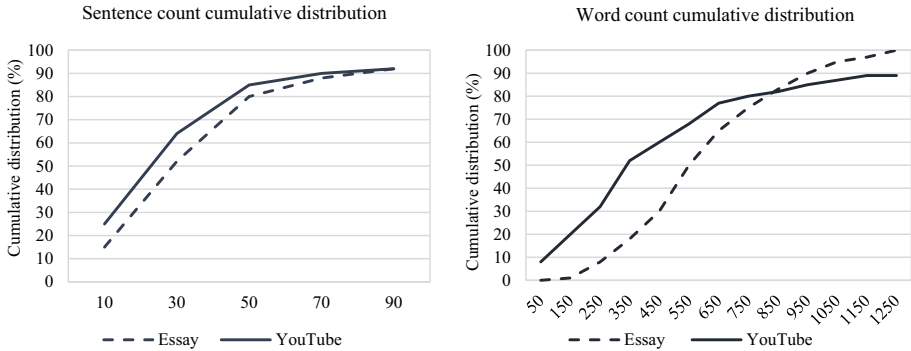


Fig. 3 Sentence and word count cumulative distribution of Essay and YouTube datasets

- The Essay dataset labels come from the authors’ questionnaire which can be considered as autognosis while the YouTube dataset labels are come from the volunteers watching the bloggers’ videos which can be treated as outer perception. As a result, employing these two datasets can help us to prove that our proposed method is valid in both cases and there is no difference whether the labels are generated by authors or other people.

4.2 Evaluation Metrics

Evaluation metrics explain the performance of a method while an important aspect of evaluation metrics is their capability to discriminate among method results. Evaluation of a deep learning based method is generally conducted by comparing the actual labels of training samples with those that are empirically labeled and choosing the best evaluation metric is highly dependent on the task. Noteworthy, the standard metric of Accuracy (Eq. 11) is used in our experiments to perform the evaluation. Where TP, TN, FP, and FN respectively refer to the true positive, true negative, false positive, and false negative.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{11}$$

We also employed MAE (Mean Absolute Error) in our experiment as the evaluation metric which was widely employed to measure the difference between the predicted score and the observed score by Big Five Inventory in APR research. It is calculated using the following equation (Eq. 12), where n shows the number of unseen instances, $\left(\begin{matrix} y_i \\ S_{x_i} \end{matrix} \right)^*$ refers to the

predicted value for the trait y_i , and $\begin{pmatrix} y_i \\ S_{x_i} \end{pmatrix}$ is the observed one. While MAE is the error measure, the lower is the better.

$$MAE = \frac{1}{n} \sum_{i=1}^n \left| \begin{pmatrix} y_i \\ S_{x_i} \end{pmatrix}^* - \begin{pmatrix} y_i \\ S_{x_i} \end{pmatrix} \right| \quad (12)$$

4.3 Experiment Description

To provide a comprehensive understanding of the efficiency of the proposed method, various experiments with several variations of the proposed method were conducted that are introduced in the following:

- CNN-AdaBoost-Rand = Random initialized word vectors are used as the input of the proposed method.
- CNN-AdaBoost-Static = Pre-trained word vectors obtained from Skip-Gram model are used as the input and their weights are not updated along the training process.
- CNN-AdaBoost-Non-Static = Pre-trained word vectors obtained from Skip-Gram model are used as the input and their weights are updated along the training process.
- CNN-AdaBoost-2channel = Combination of random initialized word vectors and pre-trained word vectors obtained from Skip-Gram model is used as the input of the proposed method.

4.4 Model Configuration and Hyper-Parameters

Due to the fact that deep neural networks require a large number of training samples to be accurately trained, a typical processor cannot be expected to perform this operation. Therefore, it is necessary to provide a powerful processor with high speed to perform the training process. All implementations of this paper were conducted on the system with Intel Xeon 2 E5-2620 2.0 GHz processor and 8 GB of RAM using Python as the programming language in the Linux environment.

The implementation process started with preprocessing the input data. To this end, the text was split into the sequence of sentences, and the period and question mark characters. Thereafter, the sentences were split into words. All letters were then reduced to lowercase all characters other than ASCII letters, exclamation marks, digits, and quotation marks were removed. Due to the fact that some documents in Essays dataset did not include periods that yielded absurdly long sentences, sentences longer than 150 words were split into sentences with 20 words (except the last piece that could be shorter).

In the following, in order to be able to use words as the input of the proposed method, they must be converted to vectors. In this regard, we trained Skip-Gram model using all existing documents while window size and word vector dimensions were respectively about 5 and 150. Notably, the learning rate of 0.025 was employed to update word vectors and minimize the loss function.

After updating the word vectors, they were fed to the convolutional layer where various filter size (3, 4, and 5) was selected and the number of filters was about 150. Rectified linear function (ReLU) was also utilized as an activation function to apply nonlinearity. It is worth mentioning that we conducted our experiments with various filter sizes and the best results were just obtained using the mentioned values.

Table 2 Configuration of the proposed method hyper-parameters

Hyper-parameters	Value
Window size	50
Word vector dimension	150
Skip-Gram learning rate	0.025
Filter size	3,4,5
Number of filters	150
Activation function	ReLU
CNN learning rate	0.01
Dropout rate	0.05
Mini-batch size	25
Number of epochs	60
Update rule	ADADELTA

The proposed neural network was trained using the training data and the output classes were determined for the multiplicity of classes by Softmax function. In order to evaluate the output of the experimental data, a cost function was adjusted and ADADELTA update rule was employed for stochastic gradient descent with a learning rate of 0.01 while mini-batch size and dropout rate were respectively about 25 and 0.05. 60 epochs were also used for training. Hyper-parameters' values used for training of the proposed method are briefly reported in Table 2.

It is worth mentioning that due to the fact that the goal of this paper is to predict the personality based on Big Five traits which include five different classes, we built five different neural networks with the same introduced structure for the five personality traits. Each network was assumed as a binary classifier to predict whether the corresponding trait is positive or negative. Moreover, all training and testing experiments were carried out based on fivefold cross-validation and the whole dataset was randomly divided into five chunks while three chunks were used as the training set and the other two chunks were used as validation and test sets. Notably, the average accuracy of each variation of the proposed method over the five fold cross-validation is reported in the result section.

4.5 Results and Analysis

As previously mentioned, we used two various datasets in our experiments and the average testing results over fivefold cross-validation obtained by different variations of the proposed method in comparison to other existing methods for the task of personality recognition on the Essay and YouTube datasets are mentioned in the following. It must be noted that while Essay dataset has been employed in other studies, we compared our obtained results with other existing methods (their results are taken from their original papers) and the results are provided in Table 3.

From the perspective of Essay dataset, all variations of the proposed method have superior performance compared to both machine learning and deep learning methods which can be due to the employment of AdaBoost algorithm that aims to separate the features by parsing the documents from different filters and boosting the classifier performance on these representations and modulate them to obtain higher performance. Notably, although AdaBoost

Table 3 Accuracy comparison of automatic classification of texts in Essay Dataset based on the big five dimensions of personality

Method	EXT	NEU	AGR	OPN	CON
TF-IDF[37]	38.13	32.96	40.35	36.24	39.11
N-Gram[37]	51.74	50.32	53.14	50.17	51.07
MairesseBaseline (MB) [5]	55.46	58.23	54.93	54.63	60.48
MairesseBaseline + CoarseAff [38]	56.45	58.33	56.03	56.73	60.68
CNN[5]	55.73	55.80	55.36	55.69	61.73
RNN[39]	56.68	56.41	55.97	56.04	62.75
CNN + Mairesse[5]	58.09	57.33	56.71	56.71	61.13
RNN + Mairesse [39]	59.73	60.23	57.81	58.27	63.44
<i>Proposed method</i>					
CNN-AdaBoost-Rand	60.05	60.91	58.11	59.34	63.94
CNN-AdaBoost-Static	60.12	61.05	58.35	59.68	64.05
CNN-AdaBoost-Non-Static	60.45	61.71	58.61	60.01	64.18
CNN-AdaBoost-2channel	61.85	62.08	59.92	60.56	64.93

algorithm is sensitive to noisy and outdated data, it is superior to most of the learning algorithms in terms of overfitting problems.

Among all variations of the proposed method, it is obvious that CNN-AdaBoost-Rand, which used random initialized word vectors as input, has the lowest accuracy while other variations that employed pre-trained word vectors as input performed slightly better which can be owing to the utilization of Skip-Gram model for providing the rich vector representation. Moreover, besides CNN-AdaBoost-Rand, CNN-AdaBoost-Static has the lowest accuracy indicating that updating weight along the training process can enhance performance. Overall, CNN-AdaBoost-2channel proved to have the highest performance for the task of personality recognition while its obtained accuracy was respectively about 61.25, 61.93, 59.02, 60.16 and 64.63% for extraversion, neuroticism, agreeableness, openness to experience, and conscientiousness.

From the perspective of YouTube dataset, it was rarely used in experiments. For this aim, we chose five well-known models as baselines to prove the efficiency of our proposed method that are listed in Table 4. The first one is the most basic method of text classification named Bayes classification based on TF-IDF features. We also compared our method with 2CNN (2-dimensions convolutional neural network), 3-CNN (3-dimensions convolutional neural network), 1-LSTM (Single RNN), 2-LSTM (bi-directional long short term memory), and 2-CLSTM (a bidirectional LSTM concatenated with CNN).

As it is clear, all variations of the proposed method have higher accuracy compared to other methods on to have the highest performance for the task of personality recognition while its obtained accuracy was respectively about 62.11, 62.43, 60.23, 61.08 and 65.19% defines that not only our proposed method can work better on shorter sentences but also it can better detect the personality traits in the eyes of other people rather than the author himself.

Results of the experiments also demonstrated that vector representation has a great impact on the performance while updating word vectors during the training process without considering whether or not the vectors are trained in advance can also lead to achieving better results.

Table 4 Accuracy comparison of automatic classification of texts in YouTube Dataset based on the big five dimensions of personality

Method	EXT	NEU	AGR	OPN	CON
Tf-IDF + Bayes [39]	47.80	58.40	61.12	49.87	56.70
2CNN [39]	49.63	57.96	50.83	49.54	50.71
3-CNN [39]	52.08	59.94	50.46	51.33	51.85
1LSTM [39]	61.85	53.25	55.38	57.91	61.17
Bi-LSTM [39]	52.20	56.13	55.47	51.83	55.67
2CLSTM [39]	67.69	61.28	58.02	55.46	61.17
<i>Proposed method</i>					
CNN-AdaBoost-Rand	60.14	61.13	58.36	60.05	64.06
CNN-AdaBoost-Static	60.81	61.23	58.83	59.88	64.38
CNN-AdaBoost-Non-Static	60.32	62.01	58.91	60.23	64.03
CNN-AdaBoost-2channel	62.11	62.43	60.23	61.08	65.19

Table 5 Average mean absolute error (MAE) obtained from different variations of the proposed model

Model	MAE value				
	CON	OPN	AGR	NEU	EXT
CNN	0.481	0.371	0.413	0.458	0.526
CNN-AdaBoost-Rand	0.478	0.368	0.401	0.431	0.518
CNN-AdaBoost-Static	0.471	0.361	0.397	0.426	0.512
CNN-AdaBoost-Non-Static	0.435	0.357	0.394	0.417	0.493
CNN-AdaBoost-2channel	0.421	0.342	0.385	0.408	0.471

To better demonstrate the performance of variations of the proposed model, we measured the MAE value and the results are mentioned in Table 5. As can be seen, CNN-AdaBoost-2channel has the lowest MAE value which clearly confirms that not only does the proposed model have higher classification accuracy it also has the lowest mean absolute error. As can be seen, CNN-AdaBoost-2channel has the lowest MAE value which confirms its higher efficiency.

Notably, considering the fact that the training time of deep neural networks is highly related to the hardware that they are implemented on, namely modern GPUs can significantly reduce the training time, it cannot be considered as a fair measure for comparing the efficiency of deep learning based methods and it has been rarely explored as a metric for evaluation. However, it is necessary to mention that choosing an optimal model a deep learning based architecture is not possible while the definition of "optimal" is not well-defined and there is always a tradeoff between the model complexity (training and test speed) and its performance.

4.6 Ablation Study

One of the downsides of convolutional neural networks is their various hyper-parameters that must be precisely tuned to obtain an optimal model. While the values of the hyper-parameters have a great impact on the performance of the proposed model, we decided to perform an ablation study to investigate the influence of different hyper-parameters on the final classification accuracy. In this regard, we hold all setting constant and vary only one factor to examine the sensitivity of the proposed model. We report the effect of the filter size,

Table 6 The influence of the filter size on the performance of the proposed model (CNN-AdaBoost-2channel)

Filter size	Accuracy %				
	CON	OPN	AGR	NEU	EXT
(3,4,5)	64.93	60.56	59.92	62.08	61.85
(4,5,6)	63.85	59.71	59.01	61.63	60.89
(6,7,8)	62.14	59.23	58.54	60.73	59.18
(8,9,10)	61.85	58.34	57.35	59.78	58.45
(3,4,5,6)	60.18	57.45	59.87	58.26	57.83

Table 7 The influence of the filter size on the performance of the proposed model (CNN-AdaBoost-2channel)

Filter size	Accuracy %				
	CON	OPN	AGR	NEU	EXT
125	63.12	59.18	58.12	61.35	60.45
150	64.93	60.56	59.92	62.08	61.85
256	62.45	58.23	57.23	60.45	59.18
300	61.93	57.23	56.14	60.73	59.23

Table 8 The influence of the activation function on the performance of the proposed model (CNN-AdaBoost-2channel)

Filter size	Accuracy %				
	CON	OPN	AGR	NEU	EXT
ReLU	64.93	60.56	59.92	62.08	61.85
SoftMax	63.12	59.18	58.12	60.65	61.36
Tanh	63.24	59.35	57.63	59.32	60.18
SoftPlus	62.18	59.05	56.12	60.23	61.05

number of filters, activation and function on one of the variations of the proposed model (CNN-AdaBoost-2channel). The results are reported based on experiments on Essay dataset.

- To investigate the effect of the filter size, we used various filters while the other parameters were kept constant. Based on the obtained results (Table 6), different filter size has a remarkable effect on the efficiency of the model, and the best classification result is obtained when the number of filters were set to (3,4,5).
- To explore the impact of the number of filters, we only varied the number of filters while other parameters were constant. As illustrated in Table 7, it can be concluded that the number of filters has also a great effect on the efficiency of the proposed model. Moreover, the highest classification accuracy was obtained when the number of filters was set to 150.
- To study the influence of activation function, different activation functions, such as ReLU, Softmax, Tanh, SoftPlus, and linear were used in our experiments. Based on the empirical results (Table 8), the ReLU function outperformed the other activation function on all source datasets.

5 Conclusion

Personality recognition is known as one of the most interesting and practical topics in both psychology and artificial intelligence. Recently, due to the penetration of the Internet in society and the combination of human relations and artificial intelligence, personality recognition has also attracted considerable attention. Commonly, machine learning based methods have been successfully utilized for this aim. However, they are confronted with some limitations and are highly dependent on human experts for extracting appropriate features. On the other hand, by the development of deep learning, as a special type of machine learning method, they presented significant results in natural language processing, particularly personality recognition, due to their amazing capabilities in automatically extracting features and their unique structure.

In this regard, convolutional neural networks can extract low-level features from the text and have been efficiently used for text classification. It is worth mentioning that despite the considerable performance of convolutional neural networks, they are still facing some major problems. The prominent challenge of these networks is that is the features, generally n-grams, obtained from various filter sizes can play a different role in the final decision. This means that a 5-g may extract more relevant information than a 4-g about the meaning of a sentence. To fill this lacuna, we decided to give the features obtained from various filters of the convolutional neural networks to separate pooling and classification layers. Therefore, when the initial results were obtained by each of the classifiers, the AdaBoost algorithm was used to produce the overall classification results. The goal of AdaBoost algorithm is to increase the learning rate of classifiers. Our proposed method combined several weak classifiers to obtain a suitable boundary for separating data between classes, which could be used to modify classifiers that were incorrectly classified.

The performance of the proposed method was examined on two heterogeneous datasets, namely Essay and YouTube datasets. Based on the empirical results, the use of AdaBoost algorithm has efficiently increased the accuracy of classification. The proposed method was able to obtain the convergence point after 60 epochs with accuracies of 61.25%, 61.93%, 59.02%, 60.16% 64.63% for extraversion, neuroticism, agreeableness, openness to experience, and conscientiousness respectively on Essay datasets. The accuracies of the proposed method on YouTube dataset were also respectively about 62.11%, 62.43%, 60.23%, 61.08% 65.19% for extraversion, neuroticism, agreeableness, openness to experience, and conscientiousness.

Employing the multimodal learning and combination of photos, videos, and other shared content besides text for predicting individuals' personalities can be considered as possible future works. Moreover, the proposed method of this paper can be utilized in other applications of cognitive science including identifying the level of stress, anxiety, and depression.

References

1. Han S, Huang H, Tang Y (2020) "Knowledge of words: an interpretable approach for personality recognition from social media. *Knowl-Based Syst* 5:105550
2. Schultz D, Schultz SE (2015) *Psychology and work today: pearson new international edition coursesmart eTextbook*. Routledge
3. Shao Z, Song S, Jaiswal S, Shen L, Valstar M, and Gunes H (2021) "Personality recognition by modelling person-specific cognitive processes using graph representation," in proceedings of the 29th ACM international conference on multimedia, , pp. 357–366.

4. Xue D et al (2018) Deep learning-based personality recognition from text posts of online social networks. *Appl Intell* 48(11):4232–4246
5. Majumder N, Poria S, Gelbukh A, Cambria E (2017) Deep learning-based document modeling for personality detection from text. *IEEE Intell Syst* 32(2):74–79
6. Barrett HC (2020) "Towards a cognitive science of the human: cross-cultural approaches and their urgency. *Trends Cognit Sci* 2:671
7. Fraenstein ED, Flowerday S (2020) "Susceptibility to phishing on social network sites: a personality information processing model. *Computers Secur* 58:101862
8. Sadr H, Pedram MM, Teshnehlab M (2019) "A robust sentiment analysis method based on sequential combination of convolutional and recursive neural networks. *Neural Process Lett* 63:1–17
9. Sadr H, Pedram MM, Teshnehlab M (2020) Multi-view deep network: a deep model based on learning features from heterogeneous neural networks for sentiment analysis. *IEEE Access* 8:86984–86997
10. Sadr H, Pedram MM, Teshnehlab M (2021) Convolutional neural network equipped with attention mechanism and transfer learning for enhancing performance of sentiment analysis. *J AI Data Min.* <https://doi.org/10.22044/jadm.2021.9618.2100>
11. S. Yakhchi, A. Beheshti, S. M. Ghafari, and M. Orgun, "Enabling the Analysis of Personality Aspects in Recommender Systems, 2020.
12. Yang J-T, Liu G-M, and Huang SC-H (2020) "emotion transformation feature: novel feature for deception detection in videos," in 2020 IEEE international conference on image processing (ICIP), IEEE, pp. 1726–1730.
13. Nilugonda M and Madhavi K (2020) "A survey on big five personality traits prediction using tensorflow," in E3S web of conferences, vol. 184: EDP Sciences, p. 01053.
14. Jalaecian Zaferani E, Teshnehlab M, Vali M (2021) Automatic personality recognition and perception using deep learning and supervised evaluation method. *J Appl Res Indus Eng* 3:52
15. Remaida A, Abdellaoui B, Moumen A, and El Idrissi YEB (2020) "Personality traits analysis using artificial neural networks: a literature survey," in 2020 1st international conference on innovative research in applied science, engineering and technology (IRASET), IEEE, pp. 1–6.
16. Soleymanpour S, Sadr H, Soleimandarabi MN (2021) "CSCNN: Cost-sensitive convolutional neural network for encrypted traffic classification. *Neural Process Lett* 6:1–27
17. Saxena A, Khanna A, Gupta D (2020) Emotion recognition and detection methods: a comprehensive survey. *J Artif Intell Syst* 2(1):53–79
18. Castellanos HA (2016) "personality recognition applying machine learning techniques on source code metrics," in FIRE (Working Notes), pp. 25–29.
19. Salama ES, El-Khoribi RA, Shoman ME, Shalaby MAW (2021) A 3D-convolutional neural network framework with ensemble learning techniques for multi-modal emotion recognition. *Egyptian Inform J* 22(2):167–176
20. Kunte A, Panicker S (2020) "Personality prediction of social network users using ensemble and XGBoost. *Progr Comput, Anal Netw* 5:133–140
21. Baig MM, Awais MM, El-Alfy E-SM (2017) AdaBoost-based artificial neural network learning. *Neuro-computing* 248:120–126
22. Zhang B et al (2019) Ensemble learners of multiple deep CNNs for pulmonary nodules classification using CT images. *IEEE Access* 7:110358–110371
23. Golbeck J, Robles C, and Turner K (2011) "Predicting personality with social media," in CHI '11 extended abstracts on human factors in computing systems, pp. 253–262.
24. Golbeck J, Robles C, Edmondson M, and Turner K (2011) "Predicting personality from twitter," in 2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing, IEEE, pp. 149–156.
25. Quercia D, Kosinski M, Stillwell D, and Crowcroft J (2011) "Our twitter profiles, our selves: Predicting personality with twitter," in 2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing, IEEE, pp. 180–185.
26. Alam F, Stepanov EA, Riccardi G (2013) "Personality traits recognition on social network-facebook," WCPR (ICWSM-13). MA, USA, Cambridge
27. Skowron M, Tkalcic M, Ferwerda B, and Schedl M (2016) "Fusing social media cues: personality prediction from twitter and instagram," in Proceedings of the 25th international conference companion on world wide web, , pp. 107–108.
28. Li L, Li A, Hao B, Guan Z, Zhu T (2014) Predicting active users' personality based on micro-blogging behaviors. *PLoS one* 9(1):84997
29. Bai S, Zhu T, and Cheng L (2012) "Big-five personality prediction based on user behaviors at social network sites," arXiv preprint [arXiv:1204.4809](https://arxiv.org/abs/1204.4809)

30. Peng K-H, Liou L-H, Chang C-S, and Lee D-S (2015) "Predicting personality traits of Chinese users based on facebook wall posts," in 2015 24th wireless and optical communication conference (WOCC), IEEE, pp. 9–14.
31. Argamon S, Dhawle S, Koppel M, and Pennebaker JW (2005) "Lexical predictors of personality type," in proceedings of the 2005 joint annual meeting of the interface and the classification society of North America , pp. 1–16.
32. Mairesse F, Walker MA, Mehl MR, Moore RK (2007) Using linguistic cues for the automatic recognition of personality in conversation and text. *J Artif Intell Res* 30:457–500
33. Yu J and Markov K (2017) "Deep learning based personality recognition from facebook status updates," in 2017 IEEE 8th international conference on awareness science and technology (iCAST), IEEE, pp. 383–387.
34. Tandra T, Suhartono D, Wongso R, Prasetyo YL (2017) Personality prediction system from facebook users. *Procedia Computer Sci* 116:604–611
35. Wang Z, Wu C-H, Li Q-B, Yan B, Zheng K-F (2020) Encoding text information with graph convolutional networks for personality recognition. *Appl Sci* 10(12):4081
36. Mikolov T, Chen K, Corrado G, and Dean J (2013) "Distributed representations of words and phrases and their compositionality, Nips,".
37. Xue D et al (2017) Personality recognition on social media with label distribution learning. *IEEE Access* 5:13478–13488
38. Mohammad SM, Kiritchenko S (2015) Using hashtags to capture fine emotion categories from tweets. *Comput Intell* 31(2):301–326
39. Sun X, Liu B, Cao J, Luo J, and Shen X (2018) "Who am I? Personality detection based on deep learning for texts," in 2018 IEEE International conference on communications (ICC), IEEE, pp. 1–6.

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