Egyptian Informatics Journal 23 (2022) 47–53

Contents lists available at ScienceDirect

Egyptian Informatics Journal

journal homepage: www.sciencedirect.com

Review

Deep learning based fusion strategies for personality prediction

Kamal El-Demerdash *, Reda A. El-Khoribi, Mahmoud A. Ismail Shoman, Sherif Abdou

Department of Information Technology, Faculty of Computers and Artificial Intelligence, Cairo University, Egypt

ARTICLE INFO

Article history:
Received 18 February 2021
Revised 6 April 2021
Accepted 8 May 2021
Available online 4 June 2021

Keywords:
NLP
Personal traits
ELMo
ULMFiT
BERT
Big five model
Classifier fusion

ABSTRACT

Automated personality trait detection from text data has emerged and gained a great deal of attention in the subject area of affective computing and sentiment analysis. Most previous work has focused on features engineering such as linguistic styles and psycholinguistic databases which have correlations with personality. Recently, natural language processing has been affected significantly with transfer learning based on feature extraction and fine-tuning pre-trained language models. We propose a new deep learning-based model for personality prediction and classification using both data and classifier level fusion. The model gets benefit from, transfer learning in natural language processing through leading pre-trained language models namely Elmo, ULMFiT, and BERT. The proposed model demonstrates the powerfulness of the introduced method to be a promising personality prediction model. When evaluating the proposed method, results show a competitive and significant accuracy enhancement of about 1.25% and 3.12% in comparison to the most recent results for the two gold standard Essays and myPersonality datasets for personality detection.

Contents

* Corresponding author.
E-mail address: k.eldemerdash@grad.fci-cu.edu.eg (K. El-Demerdash).

https://doi.org/10.1016/j.eij.2021.05.004
1110-8665/

1. Introduction

Personality refers to human variations in characteristic patterns of thought, emotion, and actions [1]. It is a combination of an integration of a person's emotion, attitude, and a general features [2]. Personality characteristics and traits have a significant effect on our lives; they affect our choices for lifestyle, luxury, health, and many other desires [3]. Social networks are usually used by people to share their movements, activities, observations, and specific points of view on problems and different issues. These traces are mainly in text form, like public posts and comments, and other media such as videos and pictures. Now, at this moment while reading this paper, a great amount of text data is being generated by billions of users on the internet. This is a fertile source of information that allows us to grasp human needs, desires, and emotional situations, furthermore is considered a rich source for the identification of personality traits [4]. Data like this could offer new intelligence in personality traits detection that will be a new trend in the psychological world. Rather, these findings can directly strengthen and increase current assumptions about personality traits. In addition to that, the recent revolution in computational power has given computer models the advance than humans at predicting personal behavior. All this will affects infinite domains such as recruitment, criminal investigations, health, and well-being. Accordingly, there's a huge interest to create Natural Language Processing (NLP) models that can naturally discover an individual's personality with vital, viable applications, that can use online text data about human interest, behavior, and preferences to automatically predict individual levels of personality traits [4]. Information about personality permits us to obtain forecasts about emotions based on experiences and situations, and hence to enhance recommendation systems [5], social network analysis [6], and affective computing and sentiment analysis [7].

1.1. Research motivation

Pragmatic applications of NLP have become significantly cheaper, faster, and more accurate due to the capability of transfer learning provided by pre-trained Language Models (LM). The effectiveness of transfer learning is visible when the features obtained from the basic task are generic and can be transferred to another task. Transfer learning empowers NLP scientists to train a model on expansive text data, and after that, they easily fine-tune and adapt the model to other NLP tasks [4].

The key research concern in our paper is addressed by leveraging the state-of-the-art deep learning-based models in NLP to discover and classify the personality traits from text data regardless of data sources, linguistic styles, and psycholinguistic features used. In this paper, we have three contributions. First, we apply classifier level fusion methods to improve the overall personality classification performance, through focusing on three leading pre-trained LMs in NLP, namely the ELMO, ULMFiT, and BERT. Second, we apply data level fusion for two benchmark personality datasets to generate more reliable, precise, and helpful features than any single dataset. Third, we present a comparative study between these three famous models and report our results on the two benchmark personality datasets. The fused dataset follows the big five personality model [8] to identify and explore personal characteristics from textual data. It is a great reasonable model since it measures and calculates the personality through five dimensions including Extroversion (EXT), Neuroticism (NEU), Agreeableness (AGR), Conscientiousness (CON), and Openness (OPN). The empirical results of the proposed model across the two benchmark personality datasets indicate a substantial increase in statistical accuracy over the state-of-the-art results.

The remainder of this paper has been arranged as follows:-

In Section 2 we explain the history of the related work and present the methodology and the model architecture design in Section 3. We clarify and analyze the experimental setup and basic outcomes in Section 4. Section 5 outlines findings and possible work.

2. Related work

Affective computing and sentiment analysis have become a key and emerging branch of Artificial Intelligence (AI) [7]. Affective computing focuses on introduces novel techniques that develop and apply affective reasoning tools for emotion detection in multiple modalities and different languages. The overall objective is to create systems that interpret and adapt the human intangibles like thoughts, emotions, feelings, personality traits, moods, sentiments, and temperament that have been deemed detectable and measurable to provide intuitive and emotionally informed responses [7]. Emotions have distinct functions in decision-making [9], such as providing information on enjoyment and pain, allowing fast decisions under time pressure, concentrating attention on specific aspects of the issue, and generating dedication to decision-making. Personality traits have a significant impact on the lives of people and influence their attitudes, desires, and decisions [10]. Personality can be characterized as psychological factors that affect actions, thought, and feeling patterns of a person that distinguish characters from each other [11]. Textual data commonly exchanged on social platforms is really the most direct and reliable way for people to translate their inner thoughts and emotions into a form that others can understand [12]. In this way, social media texts will show facets of an individual's personality. Therefore, research utilized multiple empirical indicators, including linguistic cues, syntactic features, and manually or automatically generated semantic lexicons [13]. As a consequence, the use of questionnaires such as the big five inventory is no longer necessary because the linguistic markers show the personality of the individual. The personality is instead interpreted from the unstructured text. A comprehensive research survey of modern trends and evolution in the personality prediction domain has been conducted by Yash Mehta et al. [3]. With regard to textual modality, they declared that text processing, catching valuable features from texts, and adopting the right techniques are very useful steps and can yield excellent outcomes. In this context, some researchers have made a comprehensive research and found a number of useful text features such as Linguistic Inquiry and Word Count (LIWC) [14], Mairesse [15]. Nevertheless, discovering useful text features with tight connection to the user's personality is still a wide area to research [16]. Generally, features are extracted from text to fed into traditional machine learning classifiers like Support Vector Machine (SVM) [17]. Some previous research in personality prediction from plain text was achieved by James Pennebaker et al. [18] based on their popular Essays dataset. The authors collected the famous gold standard Essays Dataset, that used in wide research experiments about personality traits detection. In addition to that, they defined the correlations between the essay and personality by using LIWC features [14]. Other features were added by Francois Mairesse et al. [15] to enhance the results on the same dataset, which called Mairesse baseline. Saif Mohammad et al. [17] applied combinations of feature sets to outperform results in [15]. A new model has been proposed by Sun et al. [16] using bidirectional Long Short-Term Memory (LSTM) networks concatenated with Convolution Neural Network (CNN) was applied on the same dataset. The authors have succeeded in detecting the personality trait leveraging the features of text's structural, by introducing the idea of merge a collection of basic features for closely related sentences. Befor the era of pre-

trained LMs in NLP, Majumder et al. [1] applied deep learning CNN model on the Essays dataset, the model surpassed the base line results for all traits. Apart from the Essays dataset, Farnadi et al. [19] conducted research in personality prediction based on social media status by using the famous textual social media myPersonality dataset [20]. The authors used some features such as LIWC and Social Network Analysis in the prediction process. This dataset was used by Park et al. [21] for automatic personality detection from text using a separate regression to classify each trait. Yu and Markov [22] improved the result by using a CNN model aligned with average pooling layer on the same dataset. Yang Li et al. [23] proposed a novel multi-task learning framework based on CNN for simultaneously detecting personality traits and emotions, the authors used myPersonality dataset and built on the known correlation between personality traits and emotional behaviors. Tandera et al. [24] implemented Multi-Layer Perceptron (MLP) deep learning architecture for myPersonality dataset, this personality prediction model outperforms the average accuracy of the previous research by 70.78% for all five traits.

2.1. Transfer learning in NLP

Transfer learning is one of today's most interesting approaches in NLP. Recently, several papers introduced and discussed transfer learning using pre-trained LMs in NLP [25–27]. Applying the transfer learning technique in NLP, pushed the boundaries of understanding and producing languages forward. Therefore, the latest trend that provides the best results in the NLP domain currently, is the application of pre-trained LMs to various downstream NLP tasks. You can either fine-tune these models with your own data on a particular task or use these models to derive high-quality LM features from your text data. The state-of-the-art advances of pre-trained LMs were adapted in the personality detection domain in some previous research. The problem of personality traits classification was addressed in [4] by leveraging ULMFiT [26] technology. The authors used the Essays dataset and outperformed the average accuracy results for all traits in [1]. Moreover, the authors emphasized that there is no need to do features engineering before classification when using pre-trained LMs features. Aslan [10] used ELMo embeddings [25] along with three completely linked layers for a transcription feature-based network, to create a multimodal framework to recognize personality traits based on different modalities. The BERT [27] model outperformed the state-of-the-art accuracy results when applied on the Essays dataset by Yash Mehta et al. [28] with an average accuracy of 60.6% for all five traits. The authors developed a model that makes use of contextualized embeddings along with psycholinguistic features for personality traits prediction. However, their results show that LM features for BERT with MLP beat the conventional psycholinguistic features. In the same context, to improve the accuracy of such psychological studies, Erik Cambria et al. [29] proposed a new version of SenticNet (a commonsense knowledge base for sentiment analysis) via an ensemble of symbolic and subsymbolic AI tools. The authors used an approach to knowledge representation that is both top-down because it leverages symbolic models like semantic networks to encode meaning, and bottom-up because it uses subsymbolic methods like BERT to implicitly learn syntactic patterns from data. Another computationally efficient deep learning-based model was presented in [30], the authors fed BERT embeddings into an SVM-based ensemble method to improve the accuracy results of the Essays dataset. To the best of our knowledge, there are no previous research was conducted using a pre-trained LM with myPersonality dataset [20] in the existing literature until the moment of writing this paper.

Here we want to focus and highlight these three pre-trained LMs that used previously in the personality detection tasks. In addition to that, these models are at the core of this new trend in NLP.

2.1.1. ELMo

Words embedding is a mechanism in which separate words are symbolized in a predefined vector region as a real value vectors [31]. There are popular ways for discovering words embedding and providing them as vectors such as GloVe and Word2Vec [31,32]. They are a distributed representation or textual content that has an outstanding execution for deep learning strategies on difficult NLP problems. However, they have one disadvantage that they can't express the polysemy in different vectors. Which mean, for the same word, even if it has different meanings in the context, its vector is unchanged. This is not acceptable for the performance requirements for many NLP tasks. Embeddings from LM or ELMo [25] are created in a way to take the whole context into consideration. ELMo is created to consider the full context for the text. Specifically, it is developed to sum the weights of the interior states for deep bi-directional LM, pre-trained on a large text corpus. In resolving the issue of polysemous terms, ELMo produces better performance and outperforms the previous techniques. In addition to ELMo performance in syntactic-based relationships, it catches semantic or meaning-related relationships. Moreover, ELMo provided a significant move towards pre-training in NLP.

2.1.2. ULMFiT

The first useful approach for fine-tuning the LM is the Universal Language Model Fine-Tuning Method [26] or ULMFiT. The model proposes procedures that are important to fine-tune a LM by using the most recent stage of ASGD Weight-Dropped LSTM LM (AWD-LSTM) [33]. The model based on the same interior architecture design, without change the hyper-parameters except adapt the dropout parameters. The LM is pre-trained on a broad public-domain corpus, on Wikitext-103 consisting of 28,595 preprocessed Wikipedia articles and 103 million words. For the purpose of pre-training, the model used auto-encoder objectives and left-to-right LM. The classifier layers above the base LM encoder are simply a pooling layer (maximum and average pool) followed by a fully connected linear layer block. The authors demonstrate the importance of several novel techniques, including discriminative fine-tuning, slanted triangular learning rate, and gradual unfreezing, for retaining previous knowledge and avoiding catastrophic forgetting during fine-tuning.

2.1.3. BERT

Transformer Bidirectional Encoder Representations, or BERT [27], is a modern cutting-edge model that uses self-attention layers to understand the meaning of each word's opposite sides [34]. It is the first deeply bidirectional, unsupervised language representation from unlabeled text by jointly conditioning on both left and right context in all layers. The two major success factors are the masking of input tokens to prevent cycles where words "see themselves" indirectly and pre-training of a model of sentence relationships. BERT is also a very big model trained on a huge corpus of words. It is pre-trained with the Books Corpus (800 M words) and Wikipedia (2,500 M words). BERT is the first fine-tuning-based representation model that achieves state-of-the-art results for a range of NLP tasks, demonstrating the enormous potential of the fine-tuning method. Depending on the architecture size of the model, we have two types of pre-trained versions of BERT. BERT-Base: 12 layers (transformer blocks), 12 attention heads, and 110 million parameters and BERT-Large: 24 layers (transformer blocks), 16 attention heads, and 340 million parameters. For fine-tuning, the BERT model is first initialized with the pre-trained parameters, and all of the parameters are fine-tuned using labeled data from the downstream tasks. Each downstream task

has a separate fine-tuned model, even though they are initialized with the same pre-trained parameters. With only an additional output layer, the pre-trained BERT model can be fine-tuned to create the state-of-the-art model for the various tasks without too many task-specific architectural changes.

3. Methodology

In short, ELMo, ULMFiT, and BERT have introduced the transfer learning techniques supported by pre-trained LMs. Each model has a different mechanism, but they all work on word embedding and fine-tuning techniques. These models have considerable contributions in NLP tasks. However, determining which method is appropriate for a particular NLP task is not a simple task. The exact output depends on the specific dataset. In this context, fusion strategies can be implemented at different levels, such as data level, features level, classifiers level, and decision level. Fusion strategies could increase the average efficiency of classification and improve prediction accuracy.

Our proposed model consists of fusion strategies on data level and classifier level, to raise the overall personality prediction performance. We adopted the three pre-trained LMs, ELMo, ULMFiT, and BERT. We used the fusion of Essays and myPersonality datasets for further fine-tuning of the proposed model [35]. The objective is to increase the performance and generalize the proposed personality prediction model independent of the data source.

The high-level architecture of the proposed model for personality prediction is shown in Fig. 1. In the following subsections we will give a description of the major components of the model.

3.1. Classifier level fusion

The final outcome of fusion depends mainly on the efficiency of each classification as well as the diversity of decisions between classifiers. Actually, even the same pre-trained model with different fine-tuning parameters having different performance. Hence, it is tricky to decide which model would lead to the best performance on the dataset. Ensemble learning refers to the procedures employed to train multiple predictors and combine their outputs. The ensemble of predictors is often called a committee machine [36]. Bagging [37] and boosting [38] are two popular committee machines. But, bagging and boosting are based on the same classification algorithm, centering toward the diversity of the data samples, that appears in limitation of diversity creates when using different algorithms [39]. Also, using multiple classifiers for feature extraction is more important in NLP [40]. Thus, we pick out the fixed rules method for multiple different classifiers fusion, it is suitable for combining independent classifiers as well as computationally less expensive approach [41]. There are three assessment levels for the base classifier outputs [42], the abstract level (the classifier output is the class label for each one), the rank level (the classifier output is a ranking list of class labels), and the measurement level (the classifier output is the posterior probability of each class). The abstract and measurement level is commonly used for classifiers combination as outputs in binary tasks. The rule of combination in the abstract level is known as vote, as its name, the output class is the most class voting among all classifiers result. In the case of the measurement level, the class output calculated using the popular algebraic rules such as mean, max, and product rules. Given a n-class classification problem: $y \in \{c_1, c_2, c_3, \ldots, c_n\}$, m classifier $\{h_1, h_2, h_3, \ldots, h_m\}$ are trained in a feature space D: $\{x_1, x_2, x_3, \ldots, x_k\}$ the combination rules are defined as follows:

$$P_{mean}\{c_i|x_1, x_2, x_3, \ldots, x_k\} = \frac{1}{m}\sum_{j=1}^{m} P_{h_j}\{c_i|x_1, x_2, x_3, \ldots, x_k\} \quad (1)$$

$$P_{max}\{c_i|x_1, x_2, x_3, \ldots, x_k\} = \max_{j=1}^{m} P_{h_j}\{c_i|x_1, x_2, x_3, \ldots, x_k\} \quad (2)$$

$$P_{product}\{c_i|x_1, x_2, x_3, \ldots, x_k\} = \prod_{j=1}^{m} P_{h_j}\{c_i|x_1, x_2, x_3, \ldots, x_k\} \quad (3)$$

As we see in the previous formula, each classifier h_j has its posterior probability P_{h_j}.The prediction after fusion is given based on these fusion fixed rules. Where the posterior average probability for m classifier is the P_{mean}, the posterior maximum probability for m classifier is the P_{max}, and the posterior product probability for m classifier is the $P_{product}$. The final predictive result is 0 for the negative trait or 1 for the positive one depending on the value of posterior probability whether it is greater or less than 0.5.

3.2. Data level fusion

Data fusion is the integration of data and knowledge from several sources, in order to generate more reliable, precise, and valid results than that provided by any single data source [43]. The process of data level fusion occurs at different levels [44], there are three-level of the data fusion process, low, intermediate, and high level. Low-level data fusion combines several sources of raw data to produce new raw data. It has been noticed that the performance of the individual pre-trained LMs is biased to the length of the text used in their training [3,16,35]. Some models give better results for short text and others for long text. We thus adopt the low-level fusion to exploit the feature engineering power of the deep-learning models. This combination of data from different sources within the same domain provides us with deeper information, higher accuracy, and more specific inferences. The expectation is that the fusion of datasets from different sources enables the pre-trained LMs to train on more task features which would not be possible using any one single dataset. This in turn can yield a classifier superior to any individual classifier.

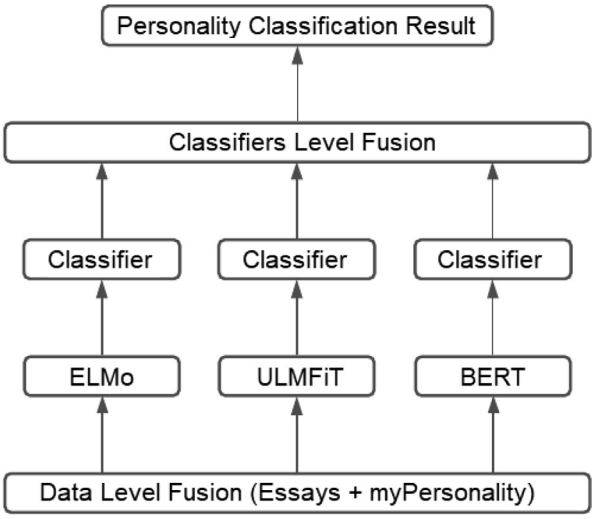

Fig. 1. High Architecture Level for the Proposed Model.

4. Experimental setup

In this section, we show the datasets used in this paper, experiments configuration, performance evaluation for each pre-trained

Table 1
The Details of the Datasets.

Name	Source	Records	Words
Essays	Students	2,467 essays	1,900,000
myPersonality	Facebook	9,917 status	143,600

Table 2
Big Five Personality Traits[4]

Five Traits	Description
Extroversion (EXT)	Is the person outgoing, talkative,and energetic versus reserved and solitary?
Neuroticism (NEU)	Is the person sensitive and nervous versus secure and confident?
Agreeableness(AGR)	Is the person trustworthy,straightforward, generous, and modest versus unreliable, complicated,meager, and boastful?
Conscientiousness (CON)	Is the person efficient and organized versus sloppy and careless?
Openness (OPN)	Is the person inventive and curious versus dogmatic and cautious?

LM at a comparative study, and the proposed model performance for personality prediction.

4.1. Dataset description

The corpus used in this paper is combined from two datasets, that are commonly used in personality prediction research, the details about the datasets are listed in Table 1. The first dataset is myPersonality dataset [20], it is Facebook status posts as raw text obtained from myPersonality project [45] sample data, myPersonality was a Facebook application that allowed its users to participate in psychological research. The application myPersonality[1] was active between 2007 and 2012 and has accumulated a large number of data that can be accessed on request.

Our research was based on a sample of the myPersonality original dataset. The second dataset is the famous Essays dataset [18]. This dataset was compiled by James Pennebaker et al. [18] to compare personality classification in text data. These valid unidentified essays were gathered from volunteers by writing stream-of-consciousness essays in a controlled environment.

Both datasets are annotated with the authors' personality traits: EXT, NEU, AGR, CON, and OPN in binary label (yes/no) values presented in Table 2, by using the author's own questionnaire of the big five model, which can be considered as autognosis [16]. The myPersonality dataset contains shorter text than the Essays dataset.

Tables 3 and 4 show the distribution of the two datasets based on personality trait type. As shown in Table 3, all traits in the Essays dataset are balanced, meanwhile in Table 4, we see the two traits NEU and OPN not balanced. Holding the original text information in order to reserve all the features could be further contributory to deep learning for personality detection as a binary classification task. Therefore, the input data is the same raw data for the three models.

4.2. Experiments configuration

We experimented with different fine-tuning techniques along with ELMo, ULMFiT, and BERT to arrive at the optimal performance for the three classifiers. We noticed similar performance across these techniques and reported the best results for the three models. First, all models were fine-tuned and tested with each dataset for comparative study purposes. Second, we fine-tuned the three models separately to the fused dataset, then each dataset predicted separately to explore how well data level fusion performs by each data source and the performance contribution for each classifier. Third, we applied the fusion methods on the classification results of the three classifiers for each dataset to obtain the final results for the proposed model.

Table 3
Distribution of Essays Dataset.

Value	EXT	NEU	AGR	CON	OPN
Yes	1276	1233	1310	1253	1217
No	1191	1234	1157	1214	1196

Table 4
Distribution of myPersonality Dataset.

Value	EXT	NEU	AGR	CON	OPN
Yes	4210	3717	5268	4556	7370
No	5707	6200	4649	5361	2547

For ELMo, we used PyTorch-based version which published by the official ELMo AllenNLP. The following action is to learn explicitly to identify personality traits from the dataset after extracting the embeddings. Contrary to all other networks, no need for an LSTM network or any other variant of Recurent Neural Networks (RNNs) since, via the bidirectional LSTM units in ELMo, the features of the word sequences are previously encoded into the embeddings. Accordingly, a few extra layers are added above this LM to train a neural regressor network that recognizes personality traits from the dataset.

For ULMFit, we used Pytorch to construct the entire model with leverage Fastai libraries to fine-tune the LM. We used backpropagation through time of 80, to allow the gradient propagation for long text sequences for LM. We used a batch size of 128 for training part, we speeded up training by a factor of 2 to 3 on GPUs. A batch size of 32 was used for the classifier since it is a bit heavier. We fine-tuned the classifier for 5 epochs as we noticed the model performance stopped improving after that.

For BERT, we used the "bert-base-uncased" model from PyTorch-pretrained-bert package for fine-tuning. As stated in the official paper of BERT, all sentences were padded or truncated to a single, fixed-length with a maximum of 512 tokens, the mini-batch as 32, the learning rate as 2e−5, and the number of training epochs as 3.

For ELMo and BERT, we added an MLP classifier for trait classification and keep the Softmax classifier in the ULMFiT model. Five independent networks were trained at every experiment using the same structure for all five personality traits. For evaluation operations, we split each dataset with 90:10 training: test ratio data respectively, we set aside a random 10% of training data to build our validation set for all models.

4.3. Results and discussion

The results will be evaluated using accuracy metric as the key performance indicator, which is the official metric used for big five personality trait prediction in the literature as a binary classification task. Accuracy is the classification percentage that is right in all data classification.

The formula for quantifying binary accuracy is:

$$Accuracy = \frac{TP + TF}{TP + TF + FP + FN} \quad (4)$$

[1] https://sites.google.com/michalkosinski.com/mypersonality/home?authuser=0

where: TP = True positive; FP = False positive; TN = True negative; FN = False negative

4.3.1. Comparative study results

All classification results obtained by using Elmo, ULMFiT, and BERT, fine-tuned and tested with each dataset separately can be seen in Tables 5 and 6. Table 5 shows the result obtained by using the Essays dataset with the three models. The fine-tuning of BERT shows the highest accuracy in EXT trait with 59.95%, CON trait with 58.93%, and OPN traits with 64.30%. While the highest accuracy in NEU trait obtained by using ELMo with 61% and the ULMFiT model outperformed in AGR trait with 59.25% accuracy. The highest average accuracy for all traits is 60.43% obtained by using BERT followed by ULMFiT with 59.88% then ELMo ranked last with 59.71%.

Table 6 shows the result obtained by using the myPersonality dataset with the three models. The highest accuracy in EXT trait with 79.50%, NEU trait with 78%, and in OPN trait with 80% is dominated by BERT. However, the highest accuracy in CON trait is 64.55% obtained by using ULMFiT, and the AGR trait outperformed by ELMo with 62.75% accuracy. The highest average accuracy for all traits is 72.10% obtained by using BERT, followed by ELMo with 71.85% then ULMFiT ranked last with 71.70%. The observations with this comparative study are, on one hand, the prediction of OPN trait, which dominated on the highest accuracy with all models across the two datasets, which indicate that it has obvious cues most easily predicted in the text data. On the other hand, the BERT model obtained the highest accuracy with EXT and OPN traits along with the highest average accuracy for all traits across the two datasets but there is no pre-trained model that dominated all big-five personality traits. Consequently, BERT has confirmed to be a great option in an industry environment where big data are available. BERT is also the more suitable choice when we have to predict a big five personality traits than ELMo or ULMFiT.

4.3.2. The proposed model results

After designing and experimented with the three models separately with each dataset and identified the optimal performance for the three classifiers, we applied them to the fused dataset (Essays training data + myPersonality training data) to find out the accuracy of classification results for each model when fine-tuned with different data sources. In addition to that, to perform the classifier fusion using the fixed rules method as shown in Section 3.1, each trait classification results are obtained separately along with each classifier through the predicted probability as 0 for the negative trait or 1 for the positive one. Then, we apply the fusion rules posterior probability for each text data point in the test data set and convert it after that to the predicted label. We noticed similar performance across these fusion fixed rules and reported the mean fusion method which was the best result for the proposed model. All classification and fusion results obtained by using Elmo, ULMFiT, and BERT, fine-tuned with fused dataset and tested with each test dataset separately can be seen in Tables 7 and 8.

Table 5
Traits Accuracy for each Model Fine-Tuning and Testing on Essays Dataset.

Model	EXT	NEU	AGR	CON	OPN	Average
ELMo	59.23	**61.00**	58.31	57.32	62.68	59.71
ULMFiT	58.89	59.92	**59.25**	57.97	63.36	59.88
BERT	**59.95**	60.16	58.80	**58.93**	**64.30**	**60.43**

Table 6
Traits Accuracy for each Model Fine-Tuning and Testing on myPersonality Dataset.

Model	EXT	NEU	AGR	CON	OPN	Average
ELMo	76.59	77.58	**62.75**	63.35	79.00	71.85
ULMFiT	77.00	76.25	62.35	**64.55**	78.35	71.70
BERT	**79.50**	**78.00**	61.00	62.00	**80.00**	**72.10**

Table 7
Traits Accuracy for Each Model Fine-tuned on Fused Dataset and Tested on Essays Dataset.

Model	EXT	NEU	AGR	CON	OPN	Average
ELMo	59.23	61.50	58.61	57.72	63.18	60.00
ULMFiT	59.59	60.29	59.25	58.47	63.26	60.17
BERT	60.85	60.75	59.80	58.90	65.30	61.10
Proposed	**61.15**	**62.20**	**60.80**	**59.52**	**65.60**	**61.85**
SOTA [28]	60.00	60.50	58.80	59.20	64.60	60.60

Table 8
Traits Accuracy for Each Model Fine-tuned on Fused Dataset and Tested on myPersonality Dataset.

Model	EXT	NEU	AGR	CON	OPN	Average
ELMo	76.59	78.00	63.30	63.75	79.60	72.25
ULMFiT	77.31	76.45	62.80	64.75	78.65	72.00
BERT	79.95	78.35	61.50	62.25	80.40	72.50
Proposed	**80.55**	79.00	**63.69**	**65.31**	**81.00**	**73.91**
SOTA [24]	78.95	**79.49**	56.52	59.62	79.31	70.78

By comparing the accuracy for each trait from Table 7 with its corresponding value in Table 5, and likewise from Table 8 with Table 6, it is found that the accuracy of most traits has improved when each pre-trained model fine-tuned on the fused dataset. However, the EXT trait with ELMo didn't have accuracy improvement in both Essays (59.23%) and mypersonality (76.59%) datasets from its corresponding in case of fused dataset, the AGR trait has the same case only in the Essays dataset with ULMFiT, the trait accuracy (59.25%) is equal in both cases of Essays dataset and fused dataset. Nevertheless, the average accuracy for all traits with each model is improved in the case of the fused dataset. From the improvement of the average accuracy results based on each model with the fused dataset shown in Tables 7 and 8, we can indicate that using more data sources increases the prediction accuracy for the LMs and, therefore, improves the results. This improvement may be different from one model to another, as it also depends on the model architecture that is used for prediction. In this context, we applied the classifier fusion methods for the three models to further improve the classification results. Based on the experimental results shown in Table 7, we can see that the proposed model dominated and outperformed all big five personality traits accuracy. The highest average accuracy for all traits is obtained by using the mean fusion method that we used at our proposed model with 61.85%. Likewise, The results in Table 8 show that the proposed model has the highest average accuracy for all traits with 73.91%. It can be seen from Table 7, and 8 that the proposed model performance is better than the performances of the separated three models, and the accuracy of the mean fusion method that we used at our proposed model along with the fused data set is the best. Moreover, we use the classes accuracy for head-to-head comparison with the state-of-the-art accuracy results, first, with respect to the Essays dataset which reported by Yash Mehta et al. [28] see Table 7, second, with respect to myPersonalty dataset by Tandera et al. [24] see Table 8.

The proposed model outperformed the average accuracy for all traits with a competitive statistical margin of about 1.25% for the gold standard Essays dataset and with a significant statistical margin of about 3.12% for the gold standard myPersonality dataset. Compared to the state-of-the-art accuracy results shown in Tables

7, and 8, the proposed model demonstrated to be a bright and powerful deep learning-based personality prediction model with respect to accuracy regardless of data source, linguistic, and psycholinguistic features used.

5. Conclusions and future work

In this paper, we introduced an efficient and explainable model for personality classification tasks based on transfer learning of pre-trained LMs features and fusion techniques. We used pre-trained LMs to train personality trait classifiers. Specifically, we leveraged the pre-trained ELMo, ULMFiT, and BERT models, we ran our experiments on two gold standard datasets.

The proposed model shows an accuracy improvement over the state-of-the-art average accuracy results for all five traits. Our results show that a fusion of sufficient data from different big five personality data sources along with fine-tuned it by the state-of-the-art pre-trained LMs as Elmo, ULMFiT, and BERT followed by a proper classifier fusion method, can outperform the previous personality prediction results without using an external feature set.

Our method performed better than the best single classifier, and also outperformed the current state-of-the-art results of the Essays dataset by 1.25% and the myPersonality dataset by 3.12%. These results indicate the maturity of transfer learning and fusion methods in personality traits detection from text. Concerning the future work we plan to experiment with the work in this paper to other affective concepts and subjectivity terms like opinions, sentiment, emotion, and mood, in order to reach to a versatile model used in social network domain.

References

[1] Majumder N, Poria S, Gelbukh A, Cambria E. Deep learning-based document modeling for personality detection from text. IEEE Intell Syst 2017;32(2):74–9.
[2] Yilmaz T, Ergil A, Ilgen B. Deep learning-based document modeling for personality detection from turkish texts. In: Proceedings of the future technologies conference (FTC).
[3] Mehta Y, Majumder N, Gelbukh A, Cambria E. Recent trends in deep learning based personality detection. Artif Intell Rev 2019. https://doi.org/10.1007/s10462-019-09770-z..
[4] El-Demerdash K, El-Khoribi RA, Shoman MAI, Abdou S. Psychological human traits detection based on universal language modeling. Egypt Inf J 2020.
[5] Lambiotte R, Kosinski M. Tracking the digital footprints of personality. In: Proceedings of the institute of electrical and electronics engineers (PIEEE). p. 1935–9.
[6] Balmaceda JM, Schiaffino S, Godoy D. How do personality traits affect communication among users in online social networks? Online Inf Rev 2014;38(1):136–53.
[7] Cambria E, Poria S, Gelbukh A, Thelwall M. Sentiment analysis is a big suitcase. IEEE Intell Syst 2017;32(6):74–80.
[8] Digman J. Personality structure: Emergence of the five-factor model. In Ann Rev Psychol 1990;41:417–40.
[9] Pfister H-R, Böhm G. The multiplicity of emotions: A framework of emotional functions in decision making. Judgm Decis Making 2008;3(1):5.
[10] Aslan S, Güdükbay U. Multimodal video-based apparent personality recognition using long short-term memory and convolutional neural networks. arXiv:1911.00381; 2019..
[11] Rr M, Op J. An introduction to the five-factor model and its applications. J Pers 1992;60(2):175–215.
[12] Xue D, WLHZ e a. Deep learning-based personality recognition from text posts of online social networks. Appl Intell 2018;48:4232–4246..
[13] Howlader Prantik et al. Predicting facebook-users' personality based on status and linguistic features via flexible regression analysis techniques. In: Proceedings of the 33rd annual ACM symposium on applied computing.
[14] Pennebaker J, Martha F, Roger B. Linguistic inquiry and word count (liwc). Lawrence Erlbaum Associates, Mahway, vol. 71.
[15] Mairesse F, Walker MA, Mehl MR, Moore RK. Using linguistic cues for the automatic recognition of personality in conversation and text. J Artif Intell Res 2007;30:457–500.
[16] Sun X, Liu B, Cao J, Luo J, Shen X. Who am i? Personality detection based on deep learning for texts. In: IEEE International Conference on Communications (ICC). IEEE. p. 1–6.
[17] Mohammad S, Kiritchenko S. Using hashtags to capture fine emotion categories from tweets. Computat Intell 2015;31(2):301–26.
[18] Pennebaker J, King LA. Linguistic styles: Language use as an individual difference. J Person Soc Psychol 1999;77(6):1296–312.
[19] Farnadi G, Zoghbi S, Moens M-F, Cock MD. How well do your facebook status updates express your personality? In: 22nd edition of the annual Belgian-Dutch conference on machine learning (BENELEARN). p. 88.
[20] Kosinski M, Matz SC, Gosling SD, Popov V, Stillwell D. Facebook as a research tool for the social sciences. In: The American psychologist. 70.543-556. DOI: 10.1037/a0039210; 2015..
[21] Park G, Schwartz HA, Eichstaedt JC, Kern ML, Kosinski M, Stillwell D, Ungar LH, Seligman MEP. Automatic personality assessment through social media language. J Personal Soc Psychol 2015;108(6):934.
[22] Yu J, Markov K. Deep learning based personality recognition from facebook status updates. In: IEEE 8th International Conference on Awareness Science and Technology (iCAST). IEEE. p. 383–7.
[23] Li Y, Kazameini A, Mehta Y, Cambria E. Multitask learning for emotion and personality detection. IEEE Trans Affect Comput 2021;1(1).
[24] Tandera T, Hendro Suhartono D, Wongso R, Prasetio YL. Personality prediction system from facebook users. In: Procedia computer science, vol. 116; 2017. p. 604–11, ISSN 1877-0509,..
[25] Peters ME, Neumann M, Iyyer M, Gardner M, Clark C, Lee K, Zettlemoyer L. Deep contextualized word representations. Assoc Comput Linguist 2018..
[26] Howard J, Ruder S. Universal language model fine-tuning for text classification. In: Proceedings of the 56th annual meeting of the association for computational linguistics (Volume 1: Long Papers). p. 328–39.
[27] Devlin J, Chang M-W, Lee K, Toutanova K. Bert:pre-training of deep bidirectional transformers for language understanding. Assoc Comput Linguist 2019..
[28] Mehta Y, Fatehi S, Kazameini A, Stachl C, Cambria E, Eetemadi S. Bottom-up and top-down: Predicting personality with psycholinguistic and language model features. ICDM 2020: 20th IEEE international conference on data mining..
[29] Cambria E, Li Y, Xing FZ, Poria S, Kwok K. Senticnet 6: Ensemble application of symbolic and subsymbolic ai for sentiment analysis. In: CIKM '20: The 29th ACM international conference on information and knowledge management.
[30] Kazameini A, Fatehi S, Mehta Y, Eetemadi S, Cambria E. Personality trait detection using bagged svm over bert word embedding ensembles. Conference: Proceedings of the fourth widening natural language processing workshop 2020. doi: https://doi.org/10.6084/m9.figshare.13012421.
[31] Pennington J, Socher R, Manning C. Glove: Global vectors for word representation In: EMNLP; 2014..
[32] Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed representations of words and phrases and their compositionality. In: Proceedings of the 26th international conference on neural information processing systems; 2013. p.3111-3119, December 05-10, Lake Tahoe, Nevada..
[33] Merity S, Keskar NS, Socher R. Regularizing and optimizing lstm language models. In: International conference on learning representations; 2018..
[34] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I. Attention is all you need. In: Advances in neural information processing systems; 2017. p. 6000–10..
[35] Sun C, Qiu X, Xu Y, Huang X. How to fine-tune BERT for text classification?. Vol. abs/1905.05583 http://arxiv.org/abs/1905.05583; 2019..
[36] Du K-L, Swamy MNS. Combining multiple learners: Data fusion and ensemble learning. In: Neural networks and statistical learning. London: Springer; 2019.
[37] Bbeiman L. Bagging predictors. Mach Learn 1990;24(2):123–40.
[38] Schapire RE. The strength of weak learnability. Mach Learn 1990;5:197–227..
[39] Liu H, Zhang L. Advancing ensemble learning performance through data transformation and classifiers fusion in granular computing context. Exp Syst Appl 2019;131:20–9.
[40] Sammons M, Christodoulopoulos C, Kordjamshidi P, Khashabi D, Srikumar V, Vijayakumar P, et al. Edison: Feature extraction for nlp, simplified. In: Proc. 10th Int. Conf. Lang. Resour. Eval.,; 2016. p. 4085–92..
[41] Hossain MA, Saddik AE, Kankanhalli M. Multimodal fusion for multimedia analysis: A survey. Multimedia Syst November 2010. doi: https://doi.org/10.1007/s00530-010-0182-0 Source: DBLP.
[42] Xu L, Krzyzak A, Suen CY. Methods of combining multiple classifiers and their applications to handwriting recognition. IEEE Trans Syst Man Cybern 1992;22(3):418–35.
[43] Rashinkar P, Krushnasamy VS. An overview of data fusion techniques. In: International conference on innovative mechanisms for industry applications (ICIMIA). p. 694–7. doi: https://doi.org/10.1109/ICIMIA.2017.7975553.
[44] Klein LA. Sensor and data fusion: A tool for information assessment and decision making. SPIE Press; 2004. p. 51.
[45] Kosinski M, Stillwell D, Graepel T. Private traits and attributes are predictable from digital records of human behavior. Nat Acad Sci 2013;110(15):5802–5. doi: https://doi.org/10.1073/pnas.1218772110.