Neurocomputing 493 (2022) 340–350

Contents lists available at ScienceDirect

Neurocomputing

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

Multitask learning for emotion and personality traits detection

Yang Li [a,*], Amirmohammad Kazemeini [b], Yash Mehta [c], Erik Cambria [d]

[a] Northwestern Polytechnical University, China
[b] Western University, Canada
[c] University College London, UK
[d] Nanyang Technological University, Singapore

ARTICLE INFO

Article history:
Received 14 March 2022
Accepted 3 April 2022
Available online 10 April 2022
Communicated by Zidong Wang

Keywords:
Personality traits detection
Emotion detection
Multitask learning
Information sharing gate

ABSTRACT

In recent years, deep learning-based automated personality traits detection has received a lot of attention, especially now, due to the massive digital footprints of an individual. Moreover, many researchers have demonstrated that there is a strong link between personality traits and emotions. In this paper, we build on the known correlation between personality traits and emotional behaviors and propose a novel transferring based multitask learning framework that simultaneously predicts both of them. We also empirically evaluate and discuss different information-sharing mechanisms between the two tasks. To ensure the high quality of the learning process, we adopt a model-agnostic meta-learning-like framework for model optimization. Our computationally efficient multitask learning model achieves the state-of-the-art performance across multiple famous personality and emotion datasets, even outperforming language model-based models.

1. Introduction

Personality traits refer to the difference among individuals in characteristic patterns of thinking, feeling, and behaving. Specifically, personality traits have been related to individual well-being, social-institutional outcomes (e.g., occupational choices, job success), and interpersonal (e.g., relationship satisfaction). In addition, in modern times, people prefer delivering their thoughts, emotions, and complaints on social media platforms, e.g., Facebook and Twitter [68]. Hence, there is a widespread interest to develop models that can use online data on human preferences and behavior (i.e., digital footprints) to automatically predict individuals' levels of personality traits for use in job screenings [45], recommender systems [44] and social network analysis. You et al. [76] found that the digital footprint on social media can be used to measure personality traits well. There are different systems in the personality traits description, and the most widely used is called the Five-Factor Model [27]. This system includes five traits that can be remembered by the acronym OCEAN: Openness (OPN), Conscientiousness (CON), Extraversion (EXT), Agreeableness (AGR), and Neuroticism (NEU). With the advancement in machine learning research and the availability of larger amounts of data, the ability and desire to detect user personality and preference are now higher than ever. While the performance of these models is not high enough to allow for the precise distinction of people based on their traits, predictions can still be 'right' on average and be utilized for digital mass persuasion [53]. However, automated personality traits prediction also raises serious concerns w.r.t. individual privacy and the conception of informed consent [52].

Recent works on this have made significant strides in machine learning-based personality traits detection [54,4,10,37,46,76,77]. However, all existing works are single-task learning in the supervised learning way. For example, Kalghatgi et al. [37] added a Multi-Layer Perceptron (MLP) over handcrafted features to do the detection, Tandera et al. [70] combined Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) together to make a better feature extraction pipeline, and finally, Mehta et al. [54] combined language models with psycholinguistic features for personality traits prediction.

As we know, emotion has a direct link to personality. According to work from [18,63,41,31], we know that neuroticism predicts higher negative emotion and lower positive affect, and conscientiousness, by contrast, is inversely associate with negative emotions, and agreeableness predicts higher positive emotion and lower negative affect, and extraversion is associated with higher positive affect and more positive subjective evaluations of daily activities, and openness is associate with a mix of positive and negative emotions [6]. Therefore, we build on our work on the result that

* Corresponding author.
E-mail addresses: liyangnpu@nwpu.edu.cn (Y. Li), akazemei@uwo.ca (A. Kazemeini), y.mehta@ucl.ac.uk (Y. Mehta), cambria@ntu.edu.sg (E. Cambria).

https://doi.org/10.1016/j.neucom.2022.04.049
0925-2312/

personality traits detection and emotion detection are complementary. However, there is no such dataset that has annotations both on personality traits and emotion. Thus, we try to conduct the multi-task learning over a separate dataset. The difficulty is how to share the information efficiently over two separate tasks. After exploring different information-sharing mechanisms e.g., Sigmoid Gate (SiG), Sigmoid weighted Linear Gate (SiLG), Across Attention Gate (CAG), and Softmax Gate (SoG) between personality traits detection and emotion detection, we propose a novel SoG based multitask learning framework SoGMTL based on CNN for simultaneously detecting personality traits and emotions (MTL is the acronyms of multitask learning), and empirical results show that our model achieves the state of the art across both personality traits and emotion datasets. Also, to ensure the high quality of the learning procedure, we propose a Model-Agnostic Meta-Learning (MAML)-like algorithm for model optimization. The main contributions of this paper are as follows:

- Proposed a multitask learning framework for personality traits detection and emotion detection.
- Designed information sharing gate SoG to share the features from two separate tasks efficiently.
- Proposed the MAML-like training algorithm to improve the performance of the multitask learning on two independent datasets.

The rest of the paper is structured as follows: Section 2 introduces related works about personality traits detection and multitask learning; Section 3 presents the proposed model and discusses several different information sharing gates; Section 4 conducts the experiments on multitask learning; Finally, Section 5 concludes the paper and presents future works.

2. Related Works

This paper mainly focuses on personality traits detection and emotion detection with a multitask learning framework. Therefore, in the related work, we will conduct a brief review of personality traits detection, emotion detection, and multitask learning, respectively.

2.1. Personality Traits Detection

It has been confirmed by researchers that online behavior is related to personality traits [34,28]. Many works have successfully applied the machine learning methods to detect the personality traits in the content generated over the social media [13,71]. Especially in the work of You et al. [76], they found that the analysis based on digital footprint was better at measuring personality traits than close others or acquaintances (friends, family, spouse, colleagues, etc.). Personality traits detection can be based on the different types of features, such as text data (self-description, social media content, etc.), demography data (gender, age, followers, etc.), and so on. One of the initial models was by Argamon et al. [4], which applied an SVM over the extracted statistical features of functional lexicons to detect the personality traits. Following this work, Farnadi et al. [23] adopted SVM to make personality traits detection over the features of network size, density, frequency of updating status, etc. Zhusupova et al. [36] detected the personality traits of Portuguese users on the Twitter platform based on demographic and social activity information. Kalghatgi et al. [37] detected the personality traits based on the neural networks (MLP) with the hand-crafted features. Su et al. [67] applied the RNN and HMM to obtain the personality traits based on the Chinese LIWC annotations extracted from the dialogue. Carducci et al. [10] also applied the SVM to do the personality traits detection, the difference between Farnadi's work [23] is that the feature they applied is text data.

Researchers also leveraged some recent developments of NLP in this field. Tandera et al. [70] made personality traits detection over the text data directly based on deep learning methods (LSTM + CNN). At the same time, Liu et al. [46] built a hierarchical structure based on Bi-RNN to learn the word and sentence representations that can infer the personality traits from three languages, i.e., English, Italian, and Spanish. Majumder et al. [49] proposed a CNN-based model to extract fixed-length features from personal documents, and then connected the learned features with 84 additional features in Mairesse's library for personality feature detection. Van et al. [73] tried to infer the personality traits based on the 275 profiles on LinkedIn, a job-related social media platform, and they concluded that extroversion could be well inferred from the self-expression of the profiles. Amirhosseini et al. [3] increased the accuracy of the MBTI dataset with Gradient boosting. Lynn et al. [47] used message-level attention instead of word-level attention to improving the result by focusing on the most relevant Facebook posts obtained from Kosinski et al. dataset [42]. Gjurkovic et al. [25] used Sentence BERT [62] to set a benchmark for their huge Reddit dataset named PANDORA, including three personality tests' labels OCEAN, MBTI, and Enneagram. Pandora tried to improve and address the deficiencies of its older version, MBTI19k [26]. Kazemeini et al. [38] fed BERT embeddings into an SVM-based ensemble method to improve the accuracy of the Essays dataset of Pennebaker et al. [58]. Finally, Mehta et al. [54] performed a thorough empirical investigation using Language Model (LM) and a variety of psycholinguistic features to identify the best combination and the impact of each feature on predicting the traits, achieving the state-of-the-art results on the Essays as well as the Kaggle MBTI dataset. Recently, Mehta et al. [55] reviewed the latest advances in deep learning-based automated personality traits prediction by focusing on effective multimodal datasets. Different from previous works, we attempt to do the personality traits detection in a new approach by designing the multitask learning framework.

2.2. Emotion Detection

Emotion detection is a merger of cognitive science and human neurology, especially when there is only text modality. Emotion detection starts with emotion models, e.g., Ekman's basic emotion model [21], Circumplex Model of Affect [65], Plutchik's wheel of emotions model [60] etc., which give a various definition about the emotion category. For example, there are six categories in Ekman's model, i.e., *anger, disgust, fear, happiness, sadness, and surprise*, however, there are eight categories in Plutchik's wheel of emotions by adding the *trust and anticipation* based on Ekman's model to form four polar-pairs of basic emotion. Therefore, emotion detection is usually treated as the multiclass classification task as there is more than one category for the emotion type. Based on these well-defined categories, approaches like lexicon based [5,69], machine learning [30,8], and deep learning [2,14,29] are applied to do the emotion detection. Especially, the advantage of deep learning has made it an efficient way to emotion detection. As it is a subtask of the sentiment analysis by predicting the emotion label, therefore, the model applied in the sentiment analysis can also be used for the emotion detection, for example, Tripto et al. [72] and Abdullah et al. [1] proposed a two-stage architecture based on LSTM and CNN to do the sentiment analysis and emotion detection at the same time. With deep learning, different contextual information can be utilized to boost emotion detection, for example, Chatterjee et al. [15] applied the LSTM to extract the contextual features from the textual dialogues and empirical results show its effectiveness in emotion detection. Similarly, Ragheb et al. [61] proposed an attention-based model to extract features from textual conversations based on BiLSTM and self-attention mechanism. Cai et al. [7] presented the multiview- and attention-based Bidirectional LSTM (Bi-LSTM) model to extract contextual features from Chinese micro-blog at multiple scales. Recently, the transformer-based model has become a popular way to do emotion detection, especially the model of Bidirectional Encoder Representations from Transformers (BERT) [17]. For example, Ratadiya et al. [51] applied the transformer-based BERT Architecture to detect the cyber abuse on the website dataset. Huang et al. [33] ensembled the hierarchical LSTM and BERT to extract the contextual features among the conversations which carried out emotion detection efficiently. Chriqui et al. [16] proposed a Hebrew BERT (HeBERT) model to do the emotion detection over the Hebrew language.

Though there are plenty of works existing in emotion detection, there are still some problems that need to cope with, one important challenge is that there is lacking massive labeled datasets to let researchers work on, to cope with this problem, we try to apply the multitask learning to do help model extract useful features from the task-related dataset.

2.3. Multitask Learning

Multitask learning has attracted lots of attention in recent years, it is to enhance performance by learning the commonalities and differences among different tasks [11,64,50]. Typically, one of the widely used models is proposed by Carunana [11,12]. And there is a shared-bottom structure across tasks. This structure reduces the risk of overfitting, but it will cause optimization conflicts due to task differences. Recently, some works attempt to resolve such conflicts by adding constraints on specific task parameters [19,56]. For example, Duong et al. [19] trained two different tasks with different data sets and added L-2 constraints between the two sets of parameters to enhance the adaptability of the model over the low resource data. Misra et al. [56] proposed the cross-stitch networks to learn a linear combination of task-specific hidden-layer embeddings. And the tasks on the semantic segmentation and surface normal prediction over the image data validate the effectiveness of this method. Yang et al. [75] generated the task-specific parameters by tensor factorization, and it achieved better performance on the task which may lead to conflicts. The drawback of this model relies on its large number of parameters which need more training data. Ma et al. [48] designed a multigate mixture-of-experts based on the work [35], it requires several experts (e.g., there are 8 experts in their settings.) to capture different features of the task, then chooses the highest gate score as the final feature. Same as the works from Yang [75], there is a large number of parameters that need more training data. Unlike previous works, this paper proposes a task-specific multitask learning framework that is used for personality traits detection and emotion detection.

3. Model Description

In this section, we will give a detailed description of our proposed model, with the framework shown in Fig. 1. Typically, there should be a shared bottom in the structure, however, it is difficult to optimize the model due to the task differences. Therefore, based on CNN, this paper designs an efficient information flow pipeline to realize the information exchange between tasks.

3.1. Preliminary

Generally, after the embedding layer, each word in the sentence is represented by the vector $x_i \in R^d$ where d is the embedding dimension. And the whole sentence is represented by $X = [x_1, x_2, \ldots, x_n]$ where n is the word number. There are three layers in the CNN model, namely the convolutional layer, the pooling layer, and the dense layer.

Same as the works in [39], we apply the convolution operation with the filter $W_k \in R^{h\times d}$ to obtain new features in the convolution layer. The operation is depicted in Eq. 1.

$$c_i = f(W_k \cdot X_{i:i+h-1} + b) \tag{1}$$

After that we can obtain a feature map with $\mathbf{c} = [c_1, c_2, \ldots, c_{n-h+1}]$. Then we apply the max-pooling layer to the feature map with $\hat{c} = \max(\mathbf{c})$ to get the most important feature that is learned with filter W_k. Finally, the dense layer is applied to map the learned features to different classes. With these three steps, we can get a powerful model in text classification.

Given two separate datasets, it is difficult to fuse features between them, to cope with such a problem, the multitask learning framework base on two separate CNNs which are connected by an information-sharing gate is proposed. As illustrated in Fig. 1, Input1 and Input2 are from two different datasets: one is labeled with personality traits and the other one is labeled with emotion. The information flow is controlled by the well-designed information sharing gate, which makes the feature sharing and transferring between two tasks without obvious conflicts, details are described in 3.2. In addition, a feature fusion-based MAML training algorithm is proposed, which can well coordinate feature selection from two different datasets. The details are described in 3.3.

In multitask learning, the gate mechanism can be applied at different levels in order to realize information sharing and fusion between different tasks. As shown in Fig. 1, the gates can be deployed after the convolutional layer, the max-pooling layer, and the dense layer respectively. Generally, information sharing over different levels will have different impacts, and these will be explained with empirical results in Section 4.3. The output of the convolutional layer is $\mathbf{c}^{ti}$, and the output of the max-pooling layer is $\hat{c}^{ti}$, where ti denotes the task index. In the next subsection, we will give a detailed discussion about gate designing and multitask training.

3.2. Information Sharing Gate

Unlike previous multi-task learning, in which different tasks are deployed on the same dataset with multiple sets of labels, each dataset has only one set of labels, and no supervision information is shared during training. We assume that the information flow between two tasks will benefit both tasks, and this has been validated in work [20]. One direct way is to design a gate mechanism to control the transfer. While there are different ways to design the gate between two tasks. Below we will give detailed descriptions about the gate designing in different ways. Before that, we need to consider the fact that the information sharing via transferring between two tasks is useful, but it is difficult to evaluate. To cope with this problem, we apply the cosine similarity between the hidden vectors after the information sharing gate to do the evaluation. Generally, the similarities between the two tasks should be neither too small nor too big, if the similarity is too small, that means there is no information flowing between the two tasks. However, the large similarity denotes all tasks sharing the same structure, which is prone to optimization conflicts. There are different ways to design the gate, and in this paper, we discussed three gate mechanisms in controlling the information flow between the tasks of

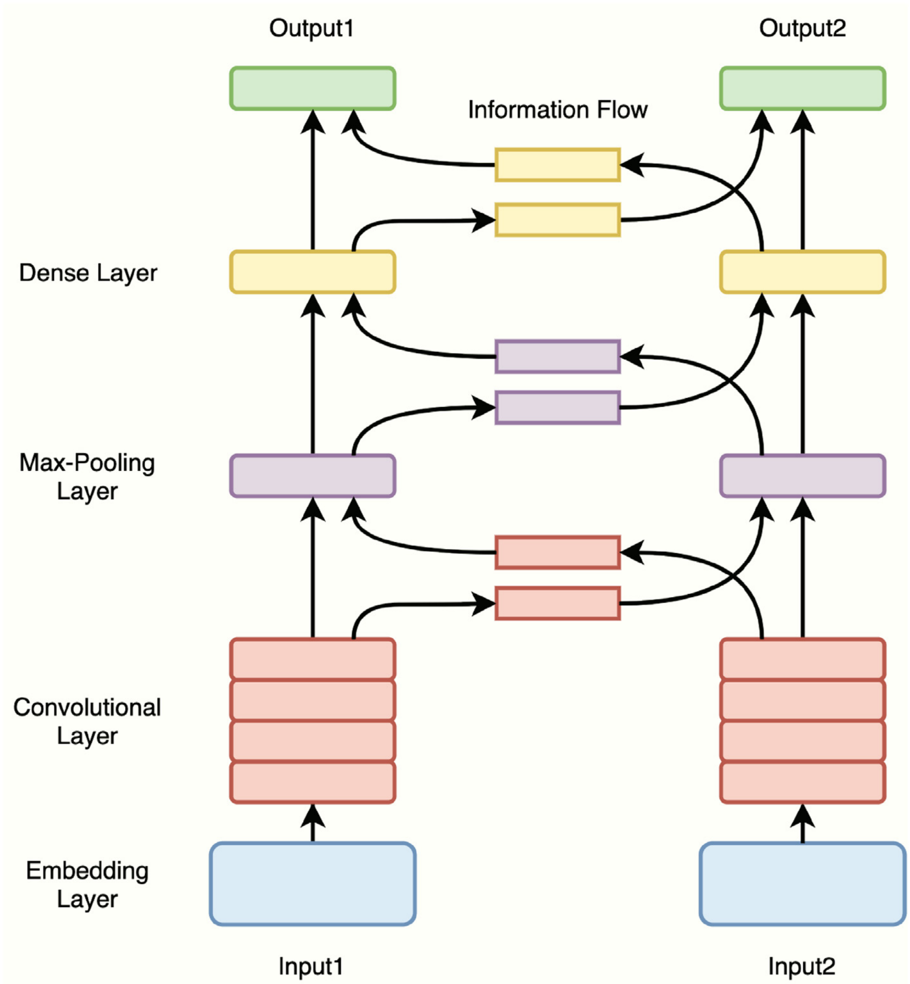


Fig. 1. The structure of the proposed framework.

emotion detection and personality traits detection. The structures are shown in Fig. 2.

SiG: The first approach is to pass the information from one network to the other one with Sigmoid function based gate directly, and it is shown in Eq. 2.

$$m_i^{t1} = \sigma(c_i^{t1})$$
$$h_i^{t2} = c_i^{t2} + m_i^{t1} \tag{2}$$

Where σ denotes the Sigmoid function on hidden value, which works as the gate between the two tasks, while $t1, t2$ are the task indexes. This gate is simple, and it lets useful and useless information pass indiscriminately between the two tasks. Therefore, it is prone to optimization conflicts, and empirical results have validated this phenomenon.

CAG: Another one is the across attention gate (CAG), which is proposed in [43]. CAG considers contextual information as the extra feature in both tasks. The procedures are described in Eq. 3.

$$m_{i,j} = c_i^{t1} W_c c_j^{t2}$$
$$\alpha_{i,j} = \frac{\exp m_{i,j}}{\sum_k \exp m_{i,k}}$$
$$h_i^{t2} = c_i^{t2} + (\sum_j \alpha_{i,j} c_j^{t2}) \tag{3}$$

where W_c denotes the model parameter. CAG has considered the features in both tasks in the cross attention calculation, and it should benefit from this cross calculation. However, the features of each task change dynamically during the optimization process, it will be difficult to get the appropriate optimal feature flow between two tasks with CAG. Also, it takes more time in the CAG calculation.

SiLG: One another gate we discussed here is the Sigmoid weighted linear gate (SiLG), which is proposed in [22], the procedures are shown in Eq. 4. This gate shows its effectiveness in reinforcement learning as an approximation of ReLU.

$$m_i^{t1} = \sigma(c_i^{t1}) \cdot c_i^{t1}$$
$$h_i^{t2} = c_i^{t2} + m_i^{t1} \tag{4}$$

While, the disadvantage of SiLG is similar with that in SiG, and it cannot select the useful information from a network by passing all information indiscriminately.

SoG: To overcome those problems, we changed the Sigmoid function in SiLG to Softmax to construct a selection gate named softmax weight gate SoG, with the result shown in the Eq. 5.

$$m_i^{t1} = softmax(c_i^{t1}) \cdot c_i^{t1}$$
$$h_i^{t2} = c_i^{t2} + m_i^{t1} \tag{5}$$

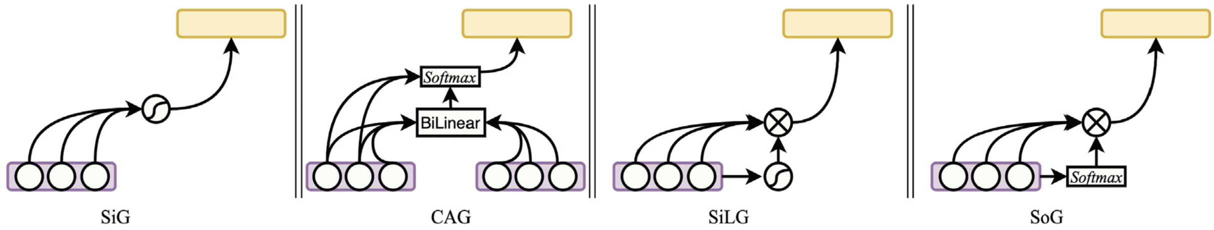


Fig. 2. The sub-net of the information flow unit.

With this change, the proposed gate mechanism is more effective: Compared with CAG, SoG has higher time efficiency in the calculation, compared with SiG and SiLG, SoG can better select features with softmax operation.

3.3. Meta Multitask Training

In multitask learning, each task has its own goal. In our framework, the two tasks are personality traits detection and emotion detection. Personality traits are always one more label, which makes personality traits detection a multi-label prediction task. Generally, it usually involves predicting one or more mutually non-exclusive class labels. Therefore, we define the objective function with the multilabel soft margin loss which is described in Eq. 6.

$$L_{Personality} = -\frac{1}{C}\sum_{i}^{C} y_i^p \cdot \log((1+\exp(-\hat{y}_i^p))^{-1}) + (1-y_i^p) \times \log(\frac{\exp(-\hat{y}_i^p)}{1+\exp(-\hat{y}_i^p)}) \tag{6}$$

where C is the class number of the personality, and it is 5 in our paper, and y_i^p stands for the label of i-th personality.

However, each sentence usually has only one emotion label, as there are multiple categories for the emotion. Thus, it is the multi-class classification task for the emotion detection by assigning each sentence to one and only one emotion label. To learn the mutilclass classification task well, we apply the cross-entropy as the objective function which is described in Eq. 7.

$$L_{Emotion} = -y^e \log(\hat{y}^e) \tag{7}$$

where y^estands for the label of the emotion. To optimize the network in a unified framework, we simply add the two loss functions together as the joint function that shows in Eq. 8,

$$L_{Multi} = L_{Personality} + L_{Emotion} \tag{8}$$

There are different ways to train the model. Generally, each task has its dataset, as we mentioned earlier, it is difficult to fuse two independent features from two random samples which are from two different data sets. Therefore, it is necessary to design an adaptive training framework for different subtasks. One tricky approach is to train the model with meta-learning, which is well adapted or generalized to new tasks and new environments that are never encountered during training. In this paper, like the k shot in MAML [24], we select the pseudo k shot (i.e., k batches shot) during training, which indicates select k more batches in one task to form the training data pair. Then we update the parameters in a MAML-like way which is depicted in Eq. 9.

$$\theta = \theta - r \cdot \nabla L_{Multi}(f(\hat{\theta})) \tag{9}$$

where θ denotes the model parameters and $\hat{\theta}$ is the adapted parameters, r is the step size, and $f(\cdot)$ stands for neural network for the multitask learning. And the algorithm we designed is shown in Algorithm 1. During the training process, we randomly select batch pairs to increase the generalization of the model.

Algorithm 1: The optimization method about MAML-like in the multitask training.

Require: Personality dataset D^p and Emotion dataset D^e
Require: Model parameters $\theta = [\theta_1, \theta_2, \ldots, \theta_n]$, and the learning rate r
1: Initialization the model parameters with random values
2: **for** select a batch from D^e **do**
3: Create a list Θ and $\mathscr{L}$ for parameter update in each step
4: Sample k batches D_k^p from D_p
5: **for** select a batch from D_k^p **do**
6: Obtain the loss L_{Multi} and its gradient descent $\nabla L_{Multi}(f_\theta)$
7: Obtain the updated parameter with $\hat{\theta} = \theta - r \cdot \nabla L_{Multi}(f(\theta))$
8: Stack $\hat{\theta}$ to Θ, and L_{Multi} to $\mathscr{L}$
9: **end for**
10: Calculate the gradient to obtain $\nabla\mathscr{L}(\theta)$ and $\nabla\Theta(\theta)$
11: Update the parameter θ with $\nabla\mathscr{L}(\theta) + \nabla\Theta(\theta)$
12: **end for**

In the Algorithm 1, line 1 is to initialize the model parameters before the training, line 2 to line 12 are the training procedure, and among those lines, line 5 to line 9 are to compute the adapted parameters based on the selected batch from Personality dataset D^p, line 10 and line 11 are to update the model parameters based on the adapted parameters.

4. Experiments

4.1. Experiment Settings

In our experiments, all the codes are written in Python with Pytorch, and they are available on Github[1]. The datasets we applied include ISEAR [66], TEC [57], and Personality [42]. There are 7,666 single-labeled sentences in ISEAR dataset contains 7 emotions, i.e., *anger, disgust, fear, joy, sadness, shame* and *guilt*. TEC includes 21,051 single-labeled sentences that are selected in tweets by pre-specified hashtags, i.e., *joy, anger, disgust, surprise, fear* and *sadness*. Personality dataset contains 9,917 multilabeled sentences in the view of the Big Five personality traits. The details about the datasets are listed in Table 1, and the distribution of the labels are shown in Fig. 3.

In the experiments, all of the datasets are split randomly in a ratio of 8:2, of which 80% are training data, and the rest are the test data. The embedding dimension is set to 300. Apart from the case of BERT which is applied with the pre-trained embedding directly,

[1] https://github.com/npuliyang/Personality-Detection-MTL

Table 1
The details of the datasets this paper involved.

Dataset	Total Number	Label Type	Label
ISEAR	7,666	Single	{anger, disgust, fear, joy, sadness, shame, guilt}
TEC	21,051	Single	{joy, anger, disgust, surprise, fear, sadness}
Personality	9,917	Multiple	{openness, conscientiousness, extraversion, agreeableness, and neuroticism}

all the embeddings are from GloVe [59]. And each document is padded to the fixed-length with 'UNK' when its length is smaller than the maximum length. All the experiments are run on CoLab with GPU support, and we run them five times to get the average results in the records. As we have mentioned, it is the multilabel prediction task for the personality traits detection, and the output of this task is a five-dimension vector after the Sigmoid function, with each value representing a trait in the Five-Factor Model. Therefore, when the value is greater than the threshold (i.e., 0.5), we then regard it as having such a trait. While, it is the multiclass classification task for emotion detection, and the output of the task is a m-dimension vector (where m is the emotion categories) after the Softmax function. Therefore, we select the index with the highest value in the m-dimension vector as the predicted label. To make a comprehensive evaluation of personality traits detection, we adopt Accuracy (ACC), Precision (P), Recall (R), and F1 as the metrics. As there are five factors in personality traits prediction i.e., the multilabel prediction task, the results of the personality traits prediction are the average of the five factors. As we can see from Fig. 3, in the ISEAR dataset, the distribution is consistent across different classes, while in the TEC dataset, about 40% data belongs to *Joy* which is the majority category. As there is no large imbalance (less 80% belongs to one class) between different classes, we adopt ACC to evaluate the emotion detection task.

4.2. Multitask Learning

Before the discussion, we need to ensure which model is the most suitable for those three datasets we selected. Therefore, we train three common used models (e.g., BERT [17], CNN [39], and LSTM [32]) on those three datasets separately, and all the results are listed in the Table 2. To maximize the performance of BERT, we follow the original work to process the input by applying the WordPiece tokenization technique, then apply the pre-trained model ('bert-base-uncased') to do personality traits detection and emotion detection respectively. For CNN, we use the classical framework proposed by Kim et al. [39] as described in Section 3.1. We apply a two-layers LSTM with the hidden dimension of 512 as another baseline. To make a fair comparison, the output of these three models are sharing the same length (i.e., 768). In the BERT, we only use the last layer as the output, and its dimension is 768. The structure of CNN is the same as in [39], with 256 filters in the size range $[3,4,5]$. And in the LSTM, the output dimension is set to 768 directly.

From Table 2, we can see that CNN has achieved the best performance in almost all datasets with different evaluation measures. That means in the personality traits detection and emotion detection, CNN has a good ability in the feature extraction. And this is also the reason why we choose CNN as the prototype in the design of the multitask model. To our surprise, BERT's performance is not the best, one possible reason is that the pre-trained model is not suitable for personality traits detection and emotion detection. In addition, the running time of BERT is too long to accept. Compared with the other two models, LSTM also has competitive results in some indicators, especially in the precision of personality traits detection, it has achieved the best performance.

There are different variants when applying information sharing gate between tasks, for example, if we apply the SiG as the gate, then the model can be named SiGMTL, where MTL stands for multitask learning. In the meantime, there also are two variants if we use a MAML-like algorithm or not, if we apply such an optimization method, then the model is named "xxx-M", otherwise, the model will be optimized by Adam [40] directly. To validate the effectiveness of the proposed SoG and proposed MAML-like algorithm, we carry out experiments with different variants and models. And experiment results are shown in Table 3 and Table 4. Specifically, the baselines we compared include:

- **SiGMTL**: It applies the SiG as the gate between two different models.
- **CAGMTL**: It applies the CAG as the gate between two different models.
- **SiLGMTL**: It applies the SiLG as the gate between two different models. And it can be treated as that proposed in work [74].
- **SoGMTL**: It is the multitask learning model we proposed by applying the SoG as the gate between two different models.

As can be seen from these two tables, SoGMTL + MAML achieves the best performance in personality traits detection under almost all conditions, except for the precision of personality traits detection that training with ISEAR. Without MAML, there is a performance decrease on SoGMTL by using the Adam algorithm directly, but it still obtains competitive results with SoG as the information sharing gate. Then it comes to the CAGMTL-M and CAGMTL, which follows the instinct that cross-attention helps

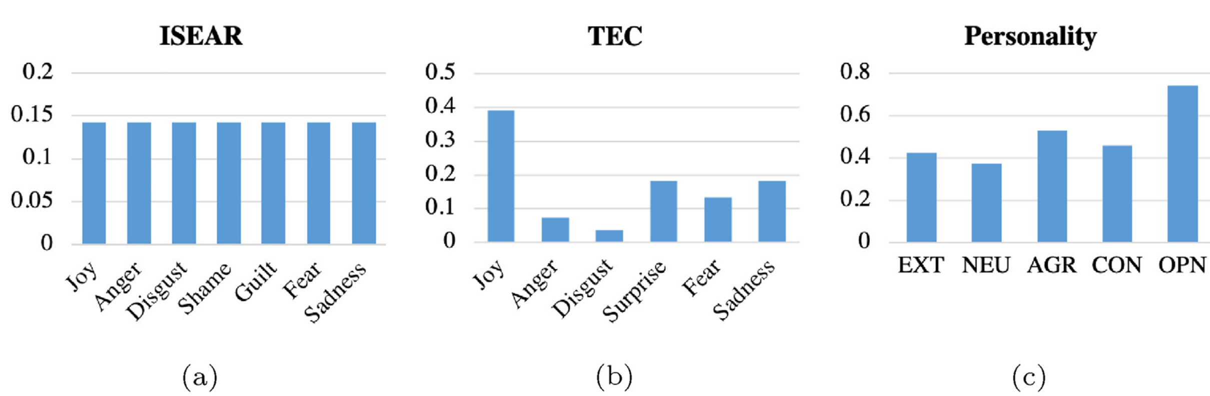


Fig. 3. The distribution of these three datasets. (a) denotes the distribution of ISEAR. (b) denotes the distribution of TEC. (b) denotes the distribution of Personality.

Table 2
The baselines results over the datasets we adopted.

Models	Personality					ISEAR	TEC
	ACC	P	R	F1	AVG	ACC	ACC
BERT	60.79%	61.53%	77.60%	68.39%	67.08%	59.28%	54.98%
LSTM	61.84%	**64.06%**	75.10%	69.13%	67.53%	54.96%	55.51%
CNN	**62.23%**	63.17%	**82.60%**	**71.54%**	**69.89%**	**59.60%**	**56.59%**

Table 3
The experiment results on the Personality + ISEAR datasets. The background color indicates the degree of the increase and decrease compared to single-task learning with CNN, with red indicating the increase and green indicating the decrease. The darker the color, the greater the degree

Models	Personality					ISEAR
	ACC	P	R	F1	Avg	ACC
SiGMTL	61.29%	63.01%	71.58%	67.61%	65.87%	57.20%
SiGMTL-M	61.70%	63.62%	74.01%	67.45%	66.70%	58.17%
CAGMTL	62.18%	64.71%	73.39%	69.08%	67.47%	58.13%
CAGMTL-M	62.41%	**64.96%**	75.27%	68.00%	67.66%	58.55%
SiLGMTL	60.49%	63.49%	68.79%	65.96%	64.68%	55.98%
SiLGMTL-M	60.81%	63.46%	70.34%	66.15%	65.19%	54.21%
SoGMTL	62.57%	64.05%	79.96%	71.09%	69.42%	59.03%
SoGMTL-M	**62.77%**	64.22%	**83.12%**	**72.45%**	**70.64%**	**59.71%**

Table 4
The experiment results on the Personality + TEC datasets. The background color indicates the degree of the increase and decrease compared to single-task learning with CNN, with red indicating the increase and green indicating the decrease. The darker the color, the greater the degree.

Models	Personality					TEC
	ACC	P	R	F1	AVG	ACC
SiGMTL	62.24%	64.59%	72.83%	68.74%	67.10%	55.45%
SiGMTL-M	61.87%	64.76%	72.79%	68.43%	66.96%	55.38%
CAGMTL	62.22%	64.23%	75.81%	69.51%	67.94%	56.49%
CAGMTL-M	62.86%	64.45%	79.29%	71.10%	69.43%	57.04%
SiLGMTL	61.54%	63.68%	72.19%	67.62%	66.26%	55.32%
SiLGMTL-M	61.21%	62.26%	77.74%	69.12%	67.58%	55.04%
SoGMTL	62.65%	63.95%	79.40%	70.82%	69.21%	56.29%
SoGMTL-M	**63.08%**	**64.79%**	**86.11%**	**73.93%**	**71.98%**	**57.15%**

the feature selection which boosts the performance. And SiLGMTL (-M) and SiGMTL(-M) come in last by achieving the worst results. Compare with single-task learning with CNN, the average improvement in personality traits detection is about 0.75% by training with ISEAR and is about 2.09% by training with TEC. And it is the same in emotion detection, the improvements of accuracy with SoGMTL + MAML is 1.32% on ISEAR, and is 0.56% on TEC respectively. However, there are performance decreases in all the other gates compared with single-task learning with CNN, and this can be seen from the green background in Table 3 and Table 4. From these observations, we can conclude that our SoGMTL has the ability to avoid the optimization conflicts between two tasks by having performance improvement compared with single-task learning. And the SoG is the best choice for the information sharing between two tasks by having the best performance among different sharing mechanisms.

When we make a horizontal comparison between TEC and ISEAR, we find that the performance on TEC dataset is generally larger than that on ISEAR. One reason may be that TEC has three times as much data as ISEAR. In terms of optimization methods, we can see that the MAML-like method is more effective than Adam as there are performance improvements in almost all cases.

To validate the computation efficiency with these three gate methods, the running time for one epoch is recorded in Fig. 4.

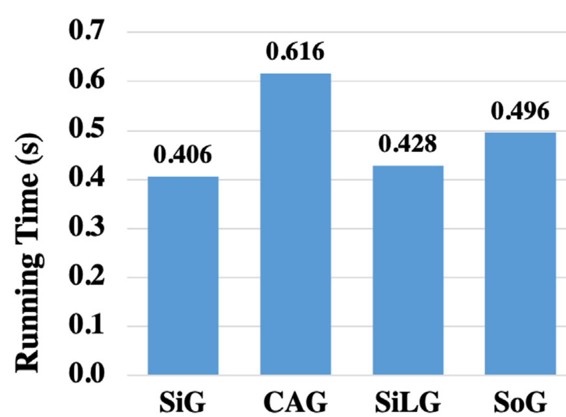


Fig. 4. The running time with different gates, y axis denotes seconds.

It can be seen from the figure that SoG is faster than CAG. As the computation of Softmax needs more complex than Sigmoid, therefore, it needs more time to compute SoG, but it still has a competitive running time compared with SiG and SiLG. Therefore, we can conclude that SoG will be one of the appropriate information-

Table 5
The performance of SoG in different levels that trained with TEC dataset, "Dense" stands for dense layer, "Max-Pooling" stands for max-pooling layer and "Conv" stands for convolutional layer.

	Personality					TEC
Settings	**ACC**	**P**	**R**	**F1**	**AVG**	**ACC**
Dense	62.45%	63.94%	79.30%	70.80%	69.12%	56.55%
Max-Pooling	62.45%	63.54%	80.82%	71.13%	69.48%	55.81%
Conv	62.90%	64.86%	81.05%	72.03%	70.21%	56.95%

Table 6
The cosine similarity between the latent vector of the two tasks in the convolutional layer and dense layer respectively when SoG is placed at different levels.

Variants	**Sim1**	**Sim2**
Dense	0.21 (-0.13)	0.47 (-0.05)
Max-Pooling	0.14 (-0.20)	0.33 (-0.19)
Conv	0.32 (-0.02)	0.47 (-0.05)
SoGMTL	0.34	0.52

sharing gates for personality traits detection and emotion detection in multitask learning.

4.3. Ablation Study

To validate the features learned by the SoG in different levels (i.e., convolutional layer, max-pooling layer, and dense layer, etc.), we place one SoG gate at different levels each a time, then see how does SoGMTL perform with a single SoG gate and the results are shown in Table 5.

From Table 5, we can see that the performance of SoG in the convolutional layer is the best, followed by the performance in the max-pooling layer. That means that the SoG in the convolutional layer is the most important in the information flow control. Furthermore, to quantify the degree of SoG's control over information flows at different locations, we calculate the cosine similarity between the latent vectors of two tasks in the convolutional layer (i.e., **Sim1**) and dense layer (i.e., **Sim2**) respectively. And the results are shown in Table 6.

SoGMTL is assumed to be the "best" status of information control in multitask learning. Therefore, it can be concluded from the table that when SoG is only placed in the dense layer, the difference between Sim1 and the best is -0.13, and the difference between Sim2 and the best is -0.05. However, both gaps become extremely small when SoG is only placed at the convolutional layer. Thus, we can know that more useful information is shared via the SoG in the convolutional layer, followed by the max-pooling layer, and less of them are shared with the dense layer. That is why the average performance is the best when SoG is only placed at the convolutional layer shown in Table 5.

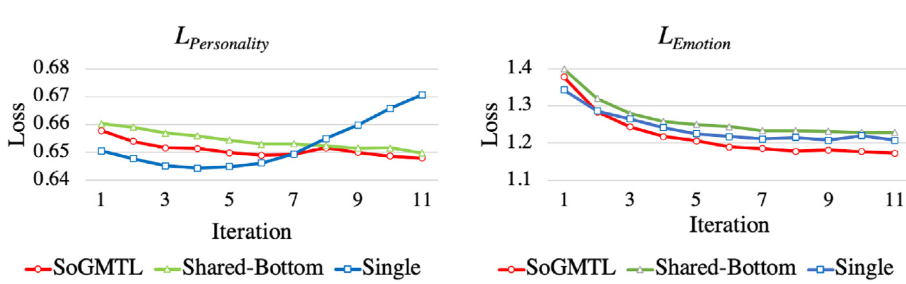
As we have mentioned before, the proposed SoGMTL has the ability to avoid optimization conflicts by using information sharing (fusion) between two tasks. To further verify such ability, we trained two tasks with two separate models as the baseline, which is the blue line named 'Single' in the Fig. 5, in this case, we optimize the two models spontaneously with L_{Multi}. We then trained the two tasks using a shared convolutional neural network at the bottom, and its result is the green line named 'Shared-Bottom' in Fig. 5. To make the comparison fair, all the parameters of these models are the same, the only difference is the way of the two networks conjunction.

From the figure, we can see that our proposed SoGMTL has the best performance than the other frameworks in emotion detection. Although its performance on personality traits detection is not the best at first, it outperformed other models after the 7th iteration. While, if there is no information sharing between two tasks, there will have a clear conflict between them. In our experiment, although the conflict between these two tasks is not apparent when using the shared bottom in the framework, it seemed to balance the conflict with performance.

4.4. Case Study

To make a clear understanding of SoGMTL's performance in multitask learning, a case study is conducted, and the results are shown in Table 7.

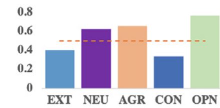
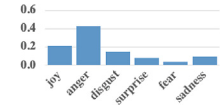
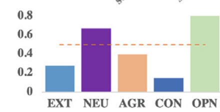
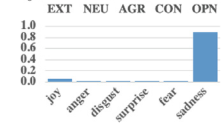
As we have mentioned earlier, personality traits detection is a multilabel classification problem, so there are multiple labels in a detection. When the value is greater than a threshold (i.e., 0.5), it indicates the user has this trait. And the cases in Table 7 show the effectiveness of SoGMTL in this task by having clear boundaries in the prediction. And emotion detection is a multiclass classification problem [9]. In the prediction, the label is determined by the maximum value. As can be seen from the image in Table 7, there is an obvious difference between the maximum value and other values. In case 1, we can see that there is an emotion consistency (both are "anger") between the two sentences from the two tasks,



(a) (b)

Fig. 5. The Loss comparisons with different frameworks. The red line denotes the loss of SoGMTL, the green line denotes the loss of shared-bottom (i.e., using the same structure in the bottom of the framework), and the blue line denotes single case where the two tasks are trained spontaneously with two separate networks by optimizing L_{Multi}.

Table 7
The case study about the SoGMTL's performance in personality traits detection and emotion detection. The red-dashed line in the predicted image indicates the threshold in the personality traits detection, it is 0.5 in our paper. *y*-axis stands for the probability of the label.

Case	Category	Sentence	True Label	Predicted
1	Personality	Damn you's a sexy bitch DAMN GIRL!!!	Neuroticism Agreeableness Openness	
	Emotion	Holding my fucking tongue	Anger	
2	Personality	Is still awake at 3:30. oh me.	Neuroticism Openness	
	Emotion	I'm home watchin this sad movie. Missing college.	Sadness	

and the prediction on the personality traits and emotion is correct with clear boundaries. However, in case 2, there is a vague emotion consistency between these two sentences from the two tasks, the emotion of the first sentence should be "anxiety" (not in the category), and the second sentence is "sadness", but there is still a clear prediction on both personality traits detection and emotion detection. From these observations, we can see that our proposed model can be well applied to personality traits detection and emotion detection in a multitask learning approach.

5. Conclusion

In this work, we designed a multitask learning framework in personality traits detection and emotion detection. In the meantime, we designed the information-sharing gate SoG between two different tasks and also discussed the difference between existing gate mechanisms. Finally, to further improve the learning performance from two different datasets, we designed a MAML-like optimization algorithm. We then conducted experiments on two emotion datasets i.e., "ISEAR" and "TEC" and one personality traits dataset "Personality", and empirical results and ablation study show the effectiveness of our proposed framework and optimization method. In future work, more contextual information like personal preference, current location, etc., will be considered.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research is supported by the National Natural Science Foundation of China (No.62103330), Fundamental Research Funds for the Central Universities (No.3102021ZDHQD09) and A*STAR under its Industry Alignment Fund (LOA Award I1901E0046).

References

[1] Malak Abdullah, Mirsad Hadzikadicy, and Samira Shaikhz. Sedat: Sentiment and emotion detection in arabic text using cnn-lstm deep learning. In 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), pages 835–840. IEEE, 2018.
[2] Fahad Mazaed Alotaibi, Classifying text-based emotions using logistic regression, VAWKUM Transactions on Computer Sciences 7 (1) (2019) 31–37.
[3] Mohammad Hossein Amirhosseini, Hassan Kazemian, Machine learning approach to personality type prediction based on the myers–briggs type indicator, Multimodal Technologies and Interaction 4 (1) (2020) 9.
[4] Shlomo Argamon, Sushant Dhawle, Moshe Koppel, James W Pennebaker, Lexical predictors of personality type, in: Proceedings of the 2005 Joint Annual Meeting of the Interface and the Classification Society of North America, 2005, pp. 1–16.
[5] Alexandra Balahur, Jesús M Hermida, and Andrés Montoyo. Detecting implicit expressions of sentiment in text based on commonsense knowledge. In Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis (WASSA 2.011), pages 53–60, 2011.
[6] Kate A. Barford, Luke D. Smillie, Openness and other big five traits in relation to dispositional mixed emotions, Personality and Individual Differences 102 (2016) 118–122.
[7] Xiaofeng Cai and Zhifeng Hao. Multi-view and attention-based bi-lstm for weibo emotion recognition. In 2018 International Conference on Network, Communication, Computer Engineering (NCCE 2018). Atlantis Press, 2018.
[8] Fabio Calefato, Filippo Lanubile, Nicole Novielli, Emotxt: a toolkit for emotion recognition from text, in: 2017 seventh international conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW), IEEE, 2017, pp. 79–80.
[9] Erik Cambria, Qian Liu, Sergio Decherchi, Frank Xing, Kenneth Kwok. SenticNet 7: A commonsense-based neurosymbolic AI framework for explainable sentiment analysis, In Proceedings of LREC (2022).
[10] Giulio Carducci, Giuseppe Rizzo, Diego Monti, Enrico Palumbo, Maurizio Morisio, Twitpersonality: Computing personality traits from tweets using word embeddings and supervised learning, Information 9 (5) (2018) 127.
[11] Rich Caruana, Multitask learning. Machine learning 28 (1) (1997) 41–75.
[12] Richard Caruana, Multitask learning: A knowledge-based source of inductive bias, in: Proceedings of the Tenth International Conference on Machine Learning, Morgan Kaufmann, 1993, pp. 41–48.
[13] Fabio Celli, Bruno Lepri, Joan-Isaac Biel, Daniel Gatica-Perez, Giuseppe Riccardi, and Fabio Pianesi. The workshop on computational personality recognition 2014. In Proceedings of the 22nd ACM international conference on Multimedia, pages 1245–1246, 2014.
[14] Ankush Chatterjee, Umang Gupta, Manoj Kumar Chinnakotla, Radhakrishnan Srikanth, Michel Galley, and Puneet Agrawal. Understanding emotions in text using deep learning and big data. Computers in Human Behavior, 93:309–317, 2019.
[15] Ankush Chatterjee, Kedhar Nath Narahari, Meghana Joshi, Puneet Agrawal, Semeval-2019 task 3: Emocontext contextual emotion detection in text, in: Proceedings of the 13th International Workshop on Semantic Evaluation, 2019, pp. 39–48.
[16] Avihay Chriqui and Inbal Yahav. Hebert & hebemo: a hebrew bert model and a tool for polarity analysis and emotion recognition. arXiv preprint arXiv:2102.01909, 2021.
[17] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.
[18] Colin G DeYoung, Lena C Quilty, and Jordan B Peterson. Between facets and domains: 10 aspects of the big five. Journal of personality and social psychology, 93(5):880, 2007.
[19] Long Duong, Trevor Cohn, Steven Bird, Paul Cook, Low resource dependency parsing: Cross-lingual parameter sharing in a neural network parser, in: Proceedings of the 53rd Annual Meeting of the Association for Computational

Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers), 2015, pp. 845–850.
[20] Christopher Ehmann, Wojciech Samek, Transferring information between neural networks, in: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2018, pp. 2361–2365.
[21] Paul Ekman, Basic emotions, Handbook of cognition and emotion 98 (45–60) (1999) 16.
[22] Stefan Elfwing, Eiji Uchibe, Kenji Doya, Sigmoid-weighted linear units for neural network function approximation in reinforcement learning, Neural Networks 107 (2018) 3–11.
[23] Golnoosh Farnadi, Susana Zoghbi, Marie-Francine Moens, Martine De Cock, How well do your facebook status updates express your personality?, in: Proceedings of the 22nd edition of the annual Belgian-Dutch conference on machine learning (BENELEARN), BNVKI-AIABN, 2013, page 88.
[24] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. arXiv preprint arXiv:1703.03400, 2017.
[25] Matej Gjurković, Mladen Karan, Iva Vukojević, Mihaela Bošnjak, and Jan Šnajder. Pandora talks: Personality and demographics on reddit. arXiv preprint arXiv:2004.04460, 2020.
[26] Matej Gjurković and Jan Šnajder. Reddit: A gold mine for personality prediction. In Proceedings of the Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media, pages 87–97, 2018.
[27] Lewis R Goldberg. The structure of phenotypic personality traits. American psychologist, 48(1):26, 1993.
[28] Samuel D Gosling, Adam A Augustine, Simine Vazire, Nicholas Holtzman, Sam Gaddis, Manifestations of personality in online social networks: Self-reported facebook-related behaviors and observable profile information, Cyberpsychology, Behavior, and Social Networking 14 (9) (2011) 483–488.
[29] Wilfredo Graterol, Jose Diaz-Amado, Yudith Cardinale, Irvin Dongo, Edmundo Lopes-Silva, Cleia Santos-Libarino, Emotion detection for social robots based on nlp transformers and an emotion ontology, Sensors 21 (4) (2021) 1322.
[30] Maryam Hasan, Elke Rundensteiner, Emmanuel Agu, Automatic emotion detection in text streams by analyzing twitter data, International Journal of Data Science and Analytics 7 (1) (2019) 35–51.
[31] Michaela Hiebler-Ragger, Jürgen Fuchshuber, Heidrun Dröscher, Christian Vajda, Andreas Fink, and Human F Unterrainer. Personality influences the relationship between primary emotions and religious/spiritual well-being. Frontiers in psychology, 9:370, 2018.
[32] Sepp Hochreiter, Jürgen Schmidhuber, Long short-term memory, Neural computation 9 (8) (1997) 1735–1780.
[33] Chenyang Huang, Amine Trabelsi, and Osmar R Zaïane. Ana at semeval-2019 task 3: Contextual emotion detection in conversations through hierarchical lstms and bert. arXiv preprint arXiv:1904.00132, 2019.
[34] David John Hughes, Moss Rowe, Mark Batey, Andrew Lee, A tale of two sites: Twitter vs. facebook and the personality predictors of social media usage, Computers in Human Behavior 28 (2) (2012) 561–569.
[35] Robert A Jacobs, Michael I Jordan, Steven J Nowlan, Geoffrey E Hinton, Adaptive mixtures of local experts, Neural computation 3 (1) (1991) 79–87.
[36] Angela Jusupova, Fernando Batista, Ricardo Ribeiro, et al., Characterizing the personality of twitter users based on their timeline information, in: Atas da 16 Conferência da Associacao Portuguesa de Sistemas de Informação, 2016, pp. 292–299.
[37] Mayuri Pundlik Kalghatgi, Manjula Ramannavar, Nandini S Sidnal, A neural network approach to personality prediction based on the big-five model, International Journal of Innovative Research in Advanced Engineering (IJIRAE) 2 (8) (2015) 56–63.
[38] Amirmohammad Kazameini, Samin Fatehi, Yash Mehta, Sauleh Eetemadi, and Erik Cambria. Personality trait detection using bagged svm over bert word embedding ensembles. In Proceedings of the The Fourth Widening Natural Language Processing Workshop. Association for Computational Linguistics, 2020.
[39] Yoon Kim. Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882, 2014.
[40] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.
[41] Emma Komulainen, Katarina Meskanen, Jari Lipsanen, Jari Marko Lahti, Pekka Jylhä, Tarja Melartin, Marieke Wichers, Erkki Isometsä, and Jesper Ekelund. The effect of personality on daily life emotional processes. PLoS One, 9(10): e110907, 2014.
[42] Michal Kosinski, David Stillwell, Thore Graepel, Private traits and attributes are predictable from digital records of human behavior, Proceedings of the national academy of sciences 110 (15) (2013) 5802–5805.
[43] Jingye Li, Meishan Zhang, Donghong Ji, and Yijiang Liu. Multi-task learning with auxiliary speaker identification for conversational emotion recognition. arXiv, pages arXiv–2003, 2020.
[44] Yang Li, Suhang Wang, Quan Pan, Haiyun Peng, Tao Yang, Erik Cambria, Learning binary codes with neural collaborative filtering for efficient recommendation systems, Knowledge-Based Systems 172 (2019) 64–75.
[45] Cynthia C.S. Liem, Markus Langer, Andrew Demetriou, Annemarie M.F. Hiemstra, Achmadnoer Sukma Wicaksana, Marise Ph Born, Cornelius J. König, Psychology meets machine learning: Interdisciplinary perspectives on algorithmic job candidate screening, in: Explainable and interpretable models in computer vision and machine learning, Springer, 2018, pp. 197–253.
[46] Fei Liu, Julien Perez, and Scott Nowson. A language-independent and compositional model for personality trait recognition from short texts. In Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers, pages 754–764, 2017.
[47] Veronica Lynn, Niranjan Balasubramanian, and H Andrew Schwartz. Hierarchical modeling for user personality prediction: The role of message-level attention. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 5306–5316, 2020.
[48] Jiaqi Ma, Zhe Zhao, Xinyang Yi, Jilin Chen, Lichan Hong, and Ed H Chi. Modeling task relationships in multi-task learning with multi-gate mixture-of-experts. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pages 1930–1939, 2018.
[49] Navonil Majumder, Soujanya Poria, Alexander Gelbukh, Erik Cambria, Deep learning-based document modeling for personality detection from text, IEEE Intelligent Systems 32 (2) (2017) 74–79.
[50] Navonil Majumder, Soujanya Poria, Haiyun Peng, Niyati Chhaya, Erik Cambria, Alexander Gelbukh, Sentiment and sarcasm classification with multitask learning, IEEE Intelligent Systems 34 (3) (2019) 38–43.
[51] Aditya Malte, Pratik Ratadiya, Multilingual cyber abuse detection using advanced transformer architecture, in: ENCON 2019–2019 IEEE Region 10 Conference (TENCON), IEEE, 2019, pp. 784–789.
[52] Sandra C Matz, Ruth E Appel, Michal Kosinski, Privacy in the age of psychological targeting, Current opinion in psychology 31 (2020) 116–121.
[53] Sandra C Matz, Michal Kosinski, Gideon Nave, and David J Stillwell. Psychological targeting as an effective approach to digital mass persuasion. Proceedings of the national academy of sciences, 114(48):12714–12719, 2017.
[54] Yash Mehta, Samin Fatehi, Amirmohammad Kazameini, Clemens Stachl, Erik Cambria, Sauleh Eetemadi, Bottom-up and top-down: Predicting personality with psycholinguistic and language model features, in: 2020 IEEE International Conference on Data Mining (ICDM), IEEE, 2020.
[55] Yash Mehta, Navonil Majumder, Alexander Gelbukh, Erik Cambria, Recent trends in deep learning based personality detection, Artificial Intelligence Review 53 (2020) 2313–2339.
[56] Ishan Misra, Abhinav Shrivastava, Abhinav Gupta, Martial Hebert, Cross-stitch networks for multi-task learning, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 3994–4003.
[57] Saif Mohammad. # emotional tweets. In * SEM 2012: The First Joint Conference on Lexical and Computational Semantics-Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012), pages 246–255, 2012.
[58] James W Pennebaker and Laura A King. Linguistic styles: Language use as an individual difference. Journal of personality and social psychology, 77(6):1296, 1999.
[59] Jeffrey Pennington, Richard Socher, Christopher D Manning, Glove: Global vectors for word representation, in: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 2014, pp. 1532–1543.
[60] Robert Plutchik, A general psychoevolutionary theory of emotion, in: Theories of emotion, Elsevier, 1980, pp. 3–33.
[61] Waleed Ragheb, Jérôme Azé, Sandra Bringay, and Maximilien Servajean. Attention-based modeling for emotion detection and classification in textual conversations. arXiv preprint arXiv:1906.07020, 2019.
[62] Nils Reimers, Iryna Gurevych, Sentence-bert: Sentence embeddings using siamese bert-networks, in: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019, pp. 3973–3983.
[63] William Revelle and Klaus R Scherer. Personality and emotion. Oxford companion to emotion and the affective sciences, 1:304–306, 2009.
[64] Sebastian Ruder. An overview of multi-task learning in deep neural networks. arXiv preprint arXiv:1706.05098, 2017.
[65] James A Russell. A circumplex model of affect. Journal of personality and social psychology, 39(6):1161, 1980.
[66] Klaus R Scherer and Harald G Wallbott. Evidence for universality and cultural variation of differential emotion response patterning. Journal of personality and social psychology, 66(2):310, 1994.
[67] Su. Ming-Hsiang, Wu. Chung-Hsien, Yu-Ting Zheng, Exploiting turn-taking temporal evolution for personality trait perception in dyadic conversations, IEEE/ACM Transactions on Audio, Speech, and Language Processing 24 (4) (2016) 733–744.
[68] Yosephine Susanto, Andrew Livingstone, Bee Chin Ng, and Erik Cambria. The hourglass model revisited. IEEE Intelligent Systems, 35(5), 96–102, 2020.
[69] Martin D Sykora, Thomas Jackson, Ann O'Brien, Suzanne Elayan, Emotive ontology: Extracting fine-grained emotions from terse, informal messages, IADIS Int. J. Comput. Sci. Inf. Syst (2013) 19–26.
[70] Tommy Tandera, Derwin Suhartono, Rini Wongso, Yen Lina Prasetio, et al. Personality prediction system from facebook users. Procedia computer science, 116:604–611, 2017.
[71] M Tkalčič, Berardina De Carolis, Marco De Gemmis, A Odić, and Andrej Košir. Preface: Empire 2014-2nd workshop on emotions and personality in personalized services. In CEUR Workshop Proceedings. CEUR-WS. org, 2014.
[72] Nafis Irtiza Tripto, Mohammed Eunus Ali, Detecting multilabel sentiment and emotions from bangla youtube comments, in: 2018 International Conference on Bangla Speech and Language Processing (ICBSLP), IEEE, 2018, pp. 1–6.
[73] Niels van de Ven, Aniek Bogaert, Alec Serlie, Mark J Brandt, Jaap JA Denissen, Personality perception based on linkedin profiles, Journal of Managerial Psychology (2017).

[74] Liqiang Xiao, Honglun Zhang, Wenqing Chen, Gated multi-task network for text classification, in: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers), 2018, pp. 726–731.
[75] Yongxin Yang and Timothy Hospedales. Deep multi-task representation learning: A tensor factorisation approach. arXiv preprint arXiv:1605.06391, 2016.
[76] Wu. Youyou, Michal Kosinski, David Stillwell, Computer-based personality judgments are more accurate than those made by humans, Proceedings of the National Academy of Sciences 112 (4) (2015) 1036–1040.
[77] A. Kazemeini, S.S. Roy, R.E. Mercer, E. Cambria, in: December). Interpretable Representation Learning for Personality Detection, In 2021 International Conference on Data Mining Workshops (ICDMW), IEEE, 2021, pp. 158–165.