EM-PERSONA: EMotion-assisted Deep Neural Framework for PERSONAlity Subtyping from Suicide Notes

Soumitra Ghosh[1], **Dhirendra Kumar Maurya**[1], **Asif Ekbal**[1] and **Pushpak Bhattacharyya**[2]
[1]Department of Computer Science and Engineering, IIT Patna, India
[2]Department of Computer Science and Engineering, IIT Bombay, India
{ghosh.soumitra2,mauryadhirendra563}@gmail.com, asif@iitp.ac.in, pb@cse.iitb.ac.in

Abstract

The World Health Organization has emphasised the need of stepping up suicide prevention efforts to meet the United Nation's Sustainable Development Goal target of 2030 (Goal 3: Good health and well-being). We address the challenging task of personality subtyping from suicide notes. Most research on personality subtyping has relied on statistical analysis and feature engineering. Moreover, state-of-the-art transformer models in the automated personality subtyping problem have received relatively less attention. We develop a novel *EMotion-assisted PERSONAlity Detection Framework (EM-PERSONA)*. We annotate the benchmark CEASE-v2.0 suicide notes dataset with personality traits across four dichotomies: *Introversion (I)-Extraversion (E), Intuition (N)-Sensing (S), Thinking (T)-Feeling (F), Judging (J)–Perceiving (P)*. Our proposed method outperforms all baselines on comprehensive evaluation using multiple state-of-the-art systems. Across the four dichotomies, *EM-PERSONA* improved accuracy by 2.04%, 3.69%, 4.52%, and 3.42%, respectively, over the highest performing single-task systems.

1 Introduction

Suicide continues to be one of the significant causes of death worldwide (Ghosh et al., 2020). Given the significance of personality as a basis for understanding psychopathology (Krueger and Tackett, 2006) and the variability in risk factors associated with suicide, subtyping patients based on their personalities can provide greater specificity than simple comparisons of suicidal to non-suicidal individuals (Ortigo et al., 2009). Pompili et al. (2008) showed that emotions such as anger, aggressiveness, anxiety, and sadness were associated with personality traits of individuals who attempted suicide.

We quote a few excerpts from suicide notes (SNs) and online personality posts (PPs) in Table 1 to show how personality discussions on public forums and genuine SNs generally follow similar language patterns. We computed cosine similarity (CosSim) scores between SNs (from CEASE-v2.0 corpus (Ghosh et al., 2022)) and PPs (from MBTI dataset[1]) and observed an alarming number of SNs having a considerable amount of word-based similarity with generic PPs. The results are shown in Table 2. We observed CosSim scores over 0.6, 0.5, 0.4 for 12, 39 and 113 SNs, respectively with respect to the PPs in MBTI dataset.

SN-1	*They just do whatever the fuck they want and justify it later.*
PP-1	*They can do anything they like to make any law they like.*
SN-2	*He is an ugly stupid faggot and we should kill him.*
PP-2	*They're oppressing you, kill them all!*

Table 1: Sample excerpts from a couple of SNs and PPs.

CosSim	>0.30	>0.40	>0.50	>0.60
# Notes	204	113	39	12

Table 2: Cosine Similarity scores between suicide notes and personality posts.

Our primary contributions are two-fold: we present a novel corpus of suicide notes annotated with personality traits across four dichotomies and develop an end-to-end multi-task emotion-assisted system for simultaneous detection of these traits from suicide notes.

2 Related Work

The Myers-Briggs Type Indicator (MBTI) (Myers, 1962), based on psychiatrist Carl Jung's ideas, is a popular personality metric that employs four dichotomies as indications of personality traits: Introversion (I) / Extraversion (E), Intuition (N) / Sensing (S), Thinking (T) / Feeling (F), Judging (J) / Perceiving (P). Another popular model like MBTI is the Big Five (Goldberg, 1993) that produces very specific and individual results, which

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

can be tedious to draw general insights and advice from test results making the practical application of the knowledge very difficult, especially when the data is scarce (as in the case of suicide note corpus). The fact that the categories for the Big Five Personality Traits are too wide and absolute to offer any meaningful insight is another issue with them. Humans are adaptive beings that adapt to their environment. In situations where we are around close friends, for instance, we could be more open, whilst in foreign settings you might be less open.

Artificial intelligence (AI) is already playing a crucial role in mental healthcare (chatbot: WoeBot, virtual assistant: Ellie (D'Alfonso, 2020)) in handling the increased demand for services, stretched workloads, high costs for treatment, and associated stigma with mental illness (Gamble, 2020). More recently, personality detection studies (Mehta et al., 2020; Yang et al., 2021; Ren et al., 2021) using computational methods have gained traction especially transformer-based (Vaswani et al., 2017) pretrained language models. However, the existing suicide note corpora (Ghosh et al., 2020, 2022) are annotated at the sentence level and existing studies do not exploit the emotional content inherent in them. This motivated us to devise an approach for utilizing the sentence-level information inherent in the existing datasets and address the closely associated tasks at the document level. Moreover, statistical analysis (Ji et al., 2021) and feature engineering (Bharadwaj et al., 2018) have been used in the bulk of the studies on this topic. However, most of the research on personality subtyping has been on domains like essays (Big Five dataset (Pennebaker and King, 1999)) and social media (MBTI Dataset) and none on the domain of suicide. This is the first attempt, to our knowledge, to identify personality subtypes of individuals who have completed suicide.

3 Dataset

We consider the benchmark CEASE-v2.0 dataset[2] (Ghosh et al., 2022) which is a fine-grained emotion annotated suicide notes corpus containing 4,932 sentences from suicide notes, each annotated independently (without any contextual information) with multi-label emotions from 15 fine-grained emotion classes.

[2]Dataset sourced from: https://www.iitp.ac.in/~ai-nlp-ml/resources.html

3.1 Personality traits annotation

Three annotators[3] sufficiently acquainted with labeling tasks and knowing the concepts of personality profile identification annotated each suicide note. For assistance in understanding the annotation task, the annotators were provided with ample instances for each personality class from the highly popular Myers Briggs Personality Type Test Dataset (MBTI dataset). The annotation task is performed across four dichotomies: *I* or *E*, *N* or *S*, *F* or *T* and *J* or *P*. For a suicide note, annotators categorised a personality trait as *Unclear (U)* if they could not evaluate the correct class owing to a lack of relevant/sufficient information. The final labels were obtained through a majority voting approach on the labels assigned by three annotators.

Traits	Distribution		κ
I-E	I: 285,	E: 71	0.58
N-S	N: 90,	S: 268	0.63
F-T	F: 238,	T: 119	0.66
J-P	J: 145,	P: 214	0.58

Table 3: Data distribution over various personality traits.

The distribution of annotated suicide notes over the various personality trait classes is shown in Table 3. As multiple raters are involved in the annotation process, we employ the Fleiss-Kappa (κ) (Spitzer et al., 1967) measure to compute the agreement among the annotators. We obtain an average Kappa agreement of 0.61 over the four personality dichotomies, indicating substantial agreement among the annotators. The score also indicates the difficulty of perceiving and synthesizing clinical ideas and labeling such tasks. The annotators relatively faced more difficulty in marking with labels *I-E* and *J-P* than *N-S* and *J-P*, which are also reflected in the attained κ scores. In both the MBTI and our annotated dataset, we observe that certain classes such as Introversion and Sensing are overrepresented while Extroversion and Intuition are relatively under-represented.

4 Methodology

Here, we discuss the proposed EMotion-assisted deep neural framework for PERSONAlity Subtyping (*EM-PERSONA*). The overall architecture is illustrated in Fig. 1.

[3]all doctoral researchers, one from the computer science discipline, two from the computational linguistics discipline

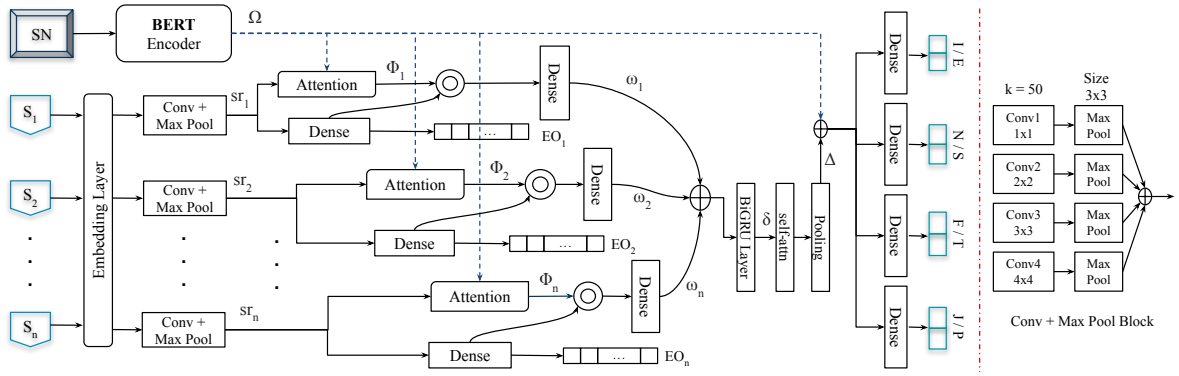

Figure 1: Architecture of the EMotion-assisted deep neural framework for PERSONAlity Subtyping.

4.1 Task Definition

Given a suicide note (N) with each sentence annotated with an emotion class[4], classify the author of the note into one of the two categories for each of the following personality dichotomies: (I/E), (N/S), (F/T), (J/P). Let $N_n^m = (s_1^m, s_2^m, .., s_n^m)$ denote a suicide note with n sentences (s) and $(e_1^m, e_2^m, ..., e_n^m)$ represents the corresponding sentence-level emotion (e) labels in the m^{th} note. The objective is to maximise the value of the following function:

$$\underset{\theta}{argmax}(\Pi_{i=0}^{m} P(y^i_{I-E}, y^i_{N-S}, y^i_{F-T}, y^i_{J-P} | s^i_n, s^i_{n-1}, ..., s^i_1; \theta)) \quad (1)$$

y: output labels, P: log likelihood function, and, θ: model parameters to be optimized.

4.2 EMotion-assisted Deep Neural Framework for PERSONAlity Subtyping (*EM-PERSONA*)

The EM-PERSONA system takes a suicide note documents as input and categorizes the author of the note into four personality classes: I/E, N/S, F/T and J/P. Each training instance comprises of a suicide note document that is encoded using the Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2019) encoder into a *contextualized document representation* (Ω). The individual sentences of the same note are processed in parallel by four convolutional and max pool layers (Conv Max Pool) of the region (k) size 1, 2, 3, and 4 and 50 filters, each of which generates sentence-level feature representations (sr_i). We use convolutional neural networks (CNN) as they are easier to parallelise, faster to train than recurrent neural networks, and effective for short sentences (Hu et al., 2018; Wang et al., 2021) (average sentence length in the CEASE corpus is 15). Word vectors, at the sentence level, are fetched from the pre-trained GloVe (Pennington et al., 2014) embedding.

To produce *contextualized sentence representations* (ϕ_i), we apply additive-attention (Bahdanau et al., 2015) between the sentence representations (sr_i) and the contextualized document representation (Ω). The attention-mechanism can be realized through the following equations:

$$\gamma = W_3^T \tanh(W_1\Omega + W_2 sr_i^c) \quad (2)$$

$$\alpha_i = \frac{exp(\gamma(\Omega sr_i^c))}{\sum_{j=1}^{c} exp(\gamma(\Omega sr_j^c))} \quad (3)$$

$$\phi_i = \sum_{t=1}^{c} \alpha_i sr_i^c \quad (4)$$

where W_1, W_2, W_3 are the learnable weight matrices, $tanh$ is a non-linear function and c is the sentence length in words.

The Conv + Max Pool outputs are also passed through sentence-specific dense layers and corresponding output layers with *softmax* activation to generate emotion classes (EO_i). The intermediate emotion-aware sentence-specific dense representations are added ($\circledcirc$) with the corresponding ϕ_i and passed through a linear layer to produce abstract emotion-aware sentence representations (ω_i).

$$\omega_i = Dense(\phi_i \circledcirc Dense_i(sr_i)) \quad (5)$$

The emotion-aware sentence representations (ω_i) are concatenated ($\oplus$) and passed through a bidirectional gated recurrent unit (BiGRU) (Cho et al., 2014) layer of 100 units to learn the contextual information. We apply multi-head self-attention (Vaswani et al., 2017) (self-attn) to attend to dif-

[4]For simplicity, we consider only the predominant emotion ($Emo1$) from CEASE-v2.0.

ferent parts of the BiGRU output and produce an *contextualized emotion-aware document representation (δ)*, which is then pooled globally.

$$\delta = BiGRU(\omega_1 \oplus \omega_2 \oplus ... \oplus \omega_n) \quad (6)$$

$$\Delta = \text{Pooling}(\text{Trans. Enc.}(\delta)) \quad (7)$$

The pre-trained BERT language model allows us to produce general contextual representations while dealing with a small supervised dataset, avoiding the need to train all the parameters from the start. We linearly concatenate Ω with the pooling layer output, Δ, and pass to four task-specific dense layers followed by the output dense layers with $softmax$ activation to get the output probability p_t^m values over the four personality trait variables.

$$p_t^m = \text{softmax}(W_t(Dense_t(\Delta \oplus \Omega)) + b_t) \quad (8)$$

W and b are learnable weight and bias matrices and t represents the four personality subtyping tasks: *I-E*, *N-S*, *F-T* and *J-P*.

where λ denotes the categorical cross-entropy loss, t represents the four personality traits tasks. α and β are the loss weights for the personality traits (PT) detection tasks and emotion recognition tasks (ER). We limit our experiments to the uniform task weighting approach, i.e., α_t and β_t are both 1.

4.3 Computation of loss

The model is trained by summing the document-level cross-entropy losses of the four personality subtasks, as well as the cross-entropy losses for the sentence-level emotion classification task.

$$\Lambda = \sum_{t=1}^{4} \alpha_t * \lambda_t^{PT} + \sum_{q=1}^{n} \beta_q * \lambda_t^{ER} \quad (9)$$

Models	F1 $^{\text{I-E}}$	F1 $^{\text{N-S}}$	F1 $^{\text{F-T}}$	F1 $^{\text{J-P}}$
Single-task baselines				
HAN	45.4	48.1	43.87	36.6
CNN+cLSTM	44.5	43	44.5	36.1
BERT	44.87	43	39.88	49.36
RoBERTa	44.46	39.88	42.69	50.58
Multitask baseline				
MT-BERT	44.35	42.68	39.90	47.31
Proposed multitask approach				
EM-PERSONA	**47.44**	**51.79**	**49.02**	**54.00**
Ablation Experiment				
EM-PERSONA$_{\text{-Emo}}$	45.53	50.27	46.96	51.40

Table 4: Scores from 10-fold cross-validation experiments. Values in bold are the maximum scores attained.

5 Experiments and Results

In this section, we discuss the experiments performed and the results and analysis.

5.1 Experimental Setup

We evaluate *EM-PERSONA* against five state-of-the-art systems: Hierarchical Attention Networks (HAN) (Yang et al., 2016), Convolutional Neural Network+Context Long Short Term Memory (CNN+cLSTM) (Poria et al., 2017), BERT-Base (Devlin et al., 2019), RoBERTa (Liu et al., 2019) and MT-BERT (Peng et al., 2020). We perform 10-fold cross-validation on the personality annotated CEASE-v2.0 dataset and consider the macro-F1 metric to evaluate our approach against multiple baselines, as class imbalance problem persists in the dataset. We discuss the details of the baselines and the hyperparameters for our experiments in Sections A.1 and A.2 of the **Appendix**.

6 Results and Discussion

Table 4 shows that the proposed *EM-PERSONA* system considerably outperforms all baseline systems, with improvements of 2.04, 3.69, 4.52, and 3.42 points over the best performing single-task systems on the four personality subtasks, respectively. The low F1 scores on the J-P trait over the HAN and CNN+cLSTM single-task baselines align with past research (Lima and de Castro, 2019; Yamada et al., 2019) where predictions on J/P dichotomy consistently underperforms compared to the other dichotomies. This is not the case for the language models, BERT and RoBERTa and the multitask systems (MT-BERT and *EM-PERSONA*), which produces comparable scores across all dichotomies, showing the effectiveness of transformer-based systems and also depicting that the correlations among the various personality traits can be effectively exploited when all the tasks are learned jointly. Commendable performance by the *EM-PERSONA* approach indicates that emotion information plays a crucial role in perceiving the personality traits of an individual through textual content-based analysis.

Ablation study: To test the impact of the emotion-assisting setup, we remove the emotion-specific dense layers in *EM-PERSONA* and see a notable drop in scores across all the personality subtasks (shown in Table 4).

Qualitative Analysis: The first example in Table 5[5] shows the effectiveness of learning the vari-

[5]Reader caution is suggested since the test cases given are

Category	*Note Excerpts*	*Actual*	MT-BERT	*EM-PERSONA*
BL & PP: FC	After many hours of thought and meditation, I have made a decision that should not be an example to anyone else ... Please tell my story on every radio and television station and in every newspaper and magazine ... to those of you who are shallow the events of this morning will be that story ... <NAME>, love you	E S F P	E S F P	E S F P
BL: PC PP: FC	.... If we had a problem it is because I loved her so much. ... we came to the understanding that for now we were not right for each other ... Unlike what has been written in the press, <NAME> and I had a great relationship for most of our lives together ... most of it ... is totally made up.	I S F J	E S F P	I S F J
BL: IC PP: PC	You have always been my soul mate and I want you to love life and know I am always with you. ... your characteristic is that of a true angel and the definition of god's love! This was the supreme Almighty's plan not mine! Look after <NAME> and <NAME> for me they are my boys you are rich. ...	E N T P	I S F J	E S F P
BL & PP: PC	Dear Mum, I am really sorry that I did this. Do not you ever think it was your fault. ... I love you so much and I could not ask for a better mum. Thank you for caring and feeding and loving me for 14 years. ... my heart cannot take this pain . I am going to miss you so much. ... I will be waiting at heaven's gates for you. ...	I S F J	E S T P	E S F P

Table 5: Sample predictions by the MT-BERT and *EM-PERSONA* systems over various categories are shown here. BL: baseline MT-BERT, PP: proposed *EM-PERSONA*, PC: partially correct, FC: fully correct, IC: fully incorrect. I: Introversion, E: Extraversion, N: Intuition, S: Sensing, T: Thinking: F: Feeling, J: Judging, P: Perceiving.

ous personality tasks jointly as both the multitasking systems, MT-BERT and *EM-PERSONA* correctly classified all the traits. In the second example, the *EM-PERSONA* system uses the emotion information in the note to classify all the traits correctly, unlike the MT-BERT system, which could only classify two personality traits correctly.

Error Analysis: The last two examples in Table 5 show some sample predictions from the MT-BERT baseline system and our proposed *EM-PERSONA* system where the models fail to classify the output classes correctly. The relevance of knowing emotion information while attempting to identify various personality traits can be realized from the observations in the third example. Here, we notice that, unlike the MT-BERT system that fails to identify a single personality trait correctly, the *EM-PERSONA* system makes correct predictions on two of the four personality traits. Rigorous analysis of instances where both the multitask systems found difficulty giving correct predictions (as in example 4) indicates that the models have a relatively more challenging time differentiating between I-E and J-P than N-S and F-T.

Test for Significance: We conducted the experiments five times and conducted a student's t-test with a 5% significance level to illustrate that the scores obtained by the proposed system have not happened by chance. We obtain the p-values of 0.039, 0.041, 0.013, and 0.009 compared with the best-performing baselines for each task, indicating that the obtained scores are statistically significant.

from genuine suicide notes and maybe deemed sensitive.

7 Conclusion

Our study focuses on artificial intelligence's assistive role, emphasising that cognitive technology is designed to enhance human intelligence rather than replace it. The proposed method is developed to serve practitioners (computer-aided diagnosis and learning) and individuals (self-monitoring) in their combined effort toward low-profile first-hand evaluation of their personalities. The findings of this study imply that (1) present state-of-the-art methods, both conversational and document encoding methods in general, fail to comprehend personality information in suicide notes to a substantial extent, (2) to improve overall system performance at the document level (such as depression, perceived burdensomeness, and thwarted belongingness), sentence-level information (such as temporal orientation, sentiment, and emotion) can be incorporated into document representations produced by existing transformer architectures, and, (3). large personality traits annotated balanced corpora are required to obtain solid findings, and the introduced resource can facilitate related studies. Identifying key subgroups of people with suicidal inclinations will help us better understand risk factors and therapies based on subtypes.

In future work, we want to address the two major limitations of our study. First, personality traits are not so simple that they can be squeezed into fixed binary categories across four dimensions, as examined in this study. Second, the short context length problem may be addressed by testing with much bigger datasets than the one used in this work.

Ethical Consideration

Our resource creation utilizes publicly available CEASE-v2.0 (Ghosh et al., 2022) benchmark suicide notes dataset. We followed the data usage restrictions and did not violate any copyright issues. This study was also evaluated and approved by our Institutional Review Board (IRB). *The data is available at* `https://www.iitp.ac.in/~ai-nlp-ml/resources.html#EMPERSONA`.

Acknowledgement

Asif Ekbal acknowledges the Young Faculty Research Fellowship (YFRF), supported by Visvesvaraya PhD scheme for Electronics and IT, Ministry of Electronics and Information Technology (MeitY), Government of India, being implemented by Digital India Corporation (formerly Media Lab Asia).

References

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Srilakshmi Bharadwaj, Srinidhi Sridhar, Rahul Choudhary, and Ramamoorthy Srinath. 2018. Persona traits identification based on myers-briggs type indicator(mbti) - A text classification approach. In *2018 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2018, Bangalore, India, September 19-22, 2018*, pages 1076–1082. IEEE.

Kyunghyun Cho, Bart van Merrienboer, Dzmitry Bahdanau, and Yoshua Bengio. 2014. On the properties of neural machine translation: Encoder-decoder approaches. In *Proceedings of SSST@EMNLP 2014, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation, Doha, Qatar, 25 October 2014*, pages 103–111.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 4171–4186.

Simon D'Alfonso. 2020. Ai in mental health. *Current Opinion in Psychology*, 36:112–117.

Alyson Gamble. 2020. Artificial intelligence and mobile apps for mental healthcare: a social informatics perspective. *Aslib J. Inf. Manag.*, 72(4):509–523.

Soumitra Ghosh, Asif Ekbal, and Pushpak Bhattacharyya. 2020. Cease, a corpus of emotion annotated suicide notes in english. In *Proceedings of The 12th Language Resources and Evaluation Conference, LREC 2020, Marseille, France, May 11-16, 2020*, pages 1618–1626.

Soumitra Ghosh, Asif Ekbal, and Pushpak Bhattacharyya. 2022. A multitask framework to detect depression, sentiment and multi-label emotion from suicide notes. *Cogn. Comput.*, 14(1):110–129.

Lewis R Goldberg. 1993. The structure of phenotypic personality traits. *American psychologist*, 48(1):26.

Yibo Hu, Yang Li, Tao Yang, and Quan Pan. 2018. Short text classification with A convolutional neural networks based method. In *15th International Conference on Control, Automation, Robotics and Vision, ICARCV 2018, Singapore, November 18-21, 2018*, pages 1432–1435. IEEE.

Shaoxiong Ji, Shirui Pan, Xue Li, Erik Cambria, Guodong Long, and Zi Huang. 2021. Suicidal ideation detection: A review of machine learning methods and applications. *IEEE Trans. Comput. Soc. Syst.*, 8(1):214–226.

Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Robert F Krueger and Jennifer L Tackett. 2006. *Personality and psychopathology*. Guilford Press.

Ana Carolina ES Lima and Leandro Nunes de Castro. 2019. Tecla: A temperament and psychological type prediction framework from twitter data. *Plos one*, 14(3):e0212844.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized BERT pretraining approach. *CoRR*, abs/1907.11692.

Yash Mehta, Samin Fatehi, Amirmohammad Kazameini, Clemens Stachl, Erik Cambria, and Sauleh Eetemadi. 2020. Bottom-up and top-down: Predicting personality with psycholinguistic and language model features. In *20th IEEE International Conference on Data Mining, ICDM 2020, Sorrento, Italy, November 17-20, 2020*, pages 1184–1189. IEEE.

Isabel Briggs Myers. 1962. The myers-briggs type indicator: Manual (1962).

Kile M Ortigo, Drew Westen, and Rebekah Bradley. 2009. Personality subtypes of suicidal adults. *The Journal of nervous and mental disease*, 197(9):687.

Yifan Peng, Qingyu Chen, and Zhiyong Lu. 2020. An empirical study of multi-task learning on BERT for biomedical text mining. In *Proceedings of the 19th SIGBioMed Workshop on Biomedical Language Processing, BioNLP 2020, Online, July 9, 2020*, pages 205–214.

James W Pennebaker and Laura A King. 1999. Linguistic styles: language use as an individual difference. *Journal of personality and social psychology*, 77(6):1296.

Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL*, pages 1532–1543.

Maurizio Pompili, Zoltán Rihmer, Hagop S Akiskal, Marco Innamorati, Paolo Iliceto, Kareen K Akiskal, David Lester, Valentina Narciso, Stefano Ferracuti, Roberto Tatarelli, et al. 2008. Temperament and personality dimensions in suicidal and nonsuicidal psychiatric inpatients. *Psychopathology*, 41(5):313–321.

Soujanya Poria, Erik Cambria, Devamanyu Hazarika, Navonil Majumder, Amir Zadeh, and Louis-Philippe Morency. 2017. Context-dependent sentiment analysis in user-generated videos. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers*, pages 873–883.

Zhancheng Ren, Qiang Shen, Xiaolei Diao, and Hao Xu. 2021. A sentiment-aware deep learning approach for personality detection from text. *Inf. Process. Manag.*, 58(3):102532.

Robert L Spitzer, Jacob Cohen, Joseph L Fleiss, and Jean Endicott. 1967. Quantification of agreement in psychiatric diagnosis: A new approach. *Archives of General Psychiatry*, 17(1):83–87.

Nitish Srivastava, Geoffrey E. Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.*, 15(1):1929–1958.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 5998–6008.

Haitao Wang, Keke Tian, Zhengjiang Wu, and Lei Wang. 2021. A short text classification method based on convolutional neural network and semantic extension. *Int. J. Comput. Intell. Syst.*, 14(1):367–375.

Kosuke Yamada, Ryohei Sasano, and Koichi Takeda. 2019. Incorporating textual information on user behavior for personality prediction. In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28 - August 2, 2019, Volume 2: Student Research Workshop*, pages 177–182.

Feifan Yang, Xiaojun Quan, Yunyi Yang, and Jianxing Yu. 2021. Multi-document transformer for personality detection. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*, pages 14221–14229.

Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alexander J. Smola, and Eduard H. Hovy. 2016. Hierarchical attention networks for document classification. In *NAACL HLT 2016, The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego California, USA, June 12-17, 2016*, pages 1480–1489.

A Appendix

A.1 Baselines

The following baseline methods are considered for the comprehensive evaluation of our proposed.

- Hierarchical Attention Network (HAN) (Yang et al., 2016): The attention mechanism in HAN takes into consideration the hierarchical structure of texts and identifies the most relevant words in a sentence and most of relevant sentences in a document while taking contextual information into account.
- CNN+cLSTM (Poria et al., 2017): A CNN is used to extract textual characteristics from utterances, after which a cLSTM is used to learn contextual information.
- BERT (Devlin et al., 2019): We experiment with the base version of the state-of-the-art BERT language model by developing four single-task binary BERT classifiers (one for each personality trait variable).
- RoBERTa (Liu et al., 2019): This is an optimized version of BERT trained with more computing power and data than BERT and is known to outperform BERT in many downstream tasks. Similar to BERT, we develop four single-task RoBERTa classifiers.
- MT-BERT (Peng et al., 2020): We build a multitask (MT) variant of BERT based on the

architecture proposed by Peng *et al.* (Peng et al., 2020) for our four personality subtypes.

A.2 Experimental Setting

We set the sequence length as 15 and the context length as 13 as the average sentence length and context length in the CEASE-v2.0 dataset. The experiments are run on an NVIDIA GeForce RTX 2080 Ti GPU. We experiment with the base version of BERT and RoBERTa imported from the Tensorflow Hub (`https://www.tensorflow.org/hub`) library. For maximum utilization of the GPU and considering the small size of the dataset, we run the MT-BERT and *EM-PERSONA* systems with a batch size of 2. Adam optimizer (Kingma and Ba, 2015) is used to train *EM-PERSONA* by minimizing the cross-entropy losses. Through grid search, we set the learning rates as 3e-5 and 2e-5 for the MT-BERT and *EM-PERSONA* systems respectively[6]. We observe empirically that setting higher epochs causes the models to overfit; hence we set the epochs as 3. We use ReLU activation on all dense layers (except the output dense) followed by a dropout (Srivastava et al., 2014) of 25% to prevent overfitting. We employ five self-attention heads for the self-attention layer, embedding dimensions = 200 and feed-forward dimensions = 400. Each task-specific dense layer has 100 neurons, whereas intermediate dense layers contain 200 neurons. To account for the non-determinism of TensorFlow GPU operations, we present F1 scores averaged across five 10-fold cross-validation runs.

[6]we experimented with epochs as 4, 6, and 8 and learning rate as 2e-5, 3e-5.