



OCEAN-AI: Open Multimodal Framework for Personality Traits Assessment and HR-Processes Automatization

Elena Ryumina¹, Dmitry Ryumin¹, Alexey Karpov²

¹St. Petersburg Federal Research Center of the Russian Academy of Sciences, St. Petersburg, Russia

²ITMO University, St. Petersburg, Russia

{ryumina.e, ryumin.d, karpov}@iiias.spb.su

Abstract

Human personality traits (PT) reflect individual differences in patterns of thinking, feeling, and behaving. Knowledge on PT may be useful in many applied tasks in our everyday life. In this paper, we present a first open-source multimodal framework called OCEAN-AI for PT assessment (PTA) and HR-processes automatization. Our framework performs PTA analyzing three modalities, including audio, video, and text, and includes three processing modules. All the modules extract heterogeneous (deep neural and hand-crafted) features and use them for a complex analysis of human's behavior. The final fourth module aggregates these six feature sets by a Siamese neural network with a gated attention mechanism. Our framework was tested on two free-available corpora, including First Impressions v2 and our MuPTA, and achieved the best results. Applying our framework, a user can automate solutions of some practical applied tasks, such as ranking potential candidates by professional responsibilities, forming efficient work teams and so on.

Index Terms: multimodal paralinguistics, analysis of speaker traits, personality traits assessment, multimodal system

1. Introduction

Human personality traits (PT) reflect individual differences in patterns of thinking, feeling, and behaving. These patterns are described by the Big Five model that is also called OCEAN [1]: Openness to experience (O), Conscientiousness (C), Extraversion (E), Agreeableness (A), and Non-neuroticism (N). Knowledge on PT may be useful in many critical tasks in everyday life, including personalized marketing, design of educational trajectories, early detection of disorders, automation of Human Resources (HR)-processes, and other tasks [2].

Automatic PT assessment (PTA) can be performed by three communication modalities: video (facial expressions, body movements, etc.), audio (energy-based, spectral and prosodic features, etc.), and text (sentiment words and their meanings, etc.). Despite the variety of proposed methods [1, 2, 3] to evaluate PT scores, existing open-source frameworks are still insufficiently developed and can not solve practical tasks. In this paper, we introduce OCEAN-AI that is the first open-source multimodal framework for PTA and HR-processes automatization.

2. OCEAN-AI Framework Architecture

OCEAN-AI is the first framework that can predict PT scores, as well as apply them to solve practical tasks, leveraging correlation coefficients between PT and target characteristics [4, 5]. OCEAN-AI offers efficient solutions for the following tasks:

- **Ranking potential candidates by professional responsibilities.** It can enhance recruitment processes and facilitating an

evaluation of potential candidates.

- **Forming efficient work teams.** Evaluating matches of potential colleagues can foster fruitful communications and minimize intra-team conflicts.
- **Predicting consumer preferences for goods.** It can enable companies to tailor their products more efficiently to consumer preferences, making them unique and personalized.

The OCEAN-AI pipeline is shown in Figure 1. The framework receives several files as an input; each file passes through four modules to predict the PT scores:

- **Video information analysis module.** The Face Mesh model¹ is used to detect a face region and facial landmarks. The Emo-Affectnet model² [6] is applied to extract deep features. Hand-crafted features such as facial asymmetry angles, distances between facial landmarks, and pairs of facial landmark coordinates are employed. Both feature sets are formed in 2 sec segments with ten frames each and are entered into two long short-term memory (LSTM) networks for extracting mid-level features (MLF).
- **Audio information analysis module.** 128 Mel filter banks with a window length of 2048 samples and a step of 512 are applied to compute log-Mel spectrograms. These spectrograms serve as an input for the pre-trained emotional VGG16 model for extracting deep features [3]. The eGeMAPS features³ are extracted as hand-crafted features. They characterize affective and physiological changes in speech and contain 25 low-level descriptors. Both sets are extracted for 2 sec segments and entered into fully connected (FCN) and LSTM networks to extract MLF, respectively.
- **Text information analysis module.** The Whisper model⁴ is employed for automatic speech recognition. The BERT model⁵ extracts deep neural features. The Linguistic Inquiry and Word Count (LIWC) set⁶ with 64 features for each word allows extracting hand-crafted features. This set is intended for a psychological analysis of English texts. Both feature sets serve as an input for two bidirectional LSTM networks with the dot-product attention mechanism to extract MLF.
- **Multimodal information fusion module.** A Siamese network with a gated attention mechanism is applied to aggregate six heterogeneous MLF and predict PT.

Thus, due to the heterogeneous nature of the features, the framework performs a complex analysis of human behavior and ac-

¹<https://developers.google.com/mediapipe>

²<https://github.com/ElenaRyumina/EMO-AffectNetModel>

³<https://audeering.github.io/opensmile-python/>

⁴<https://huggingface.co/openai/whisper-base>

⁵<https://huggingface.co/bert-base-multilingual-cased>

⁶<https://www.liwc.app/>

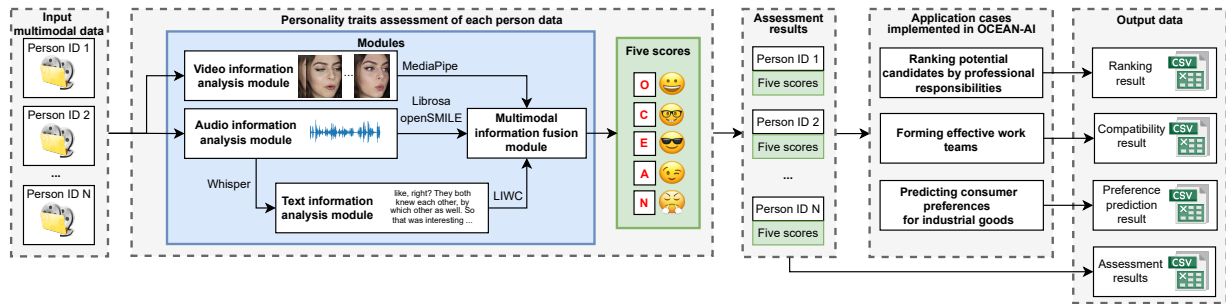


Figure 1: A pipeline of the OCEAN-AI framework.

curately evaluates PT.

All the models were trained by the TensorFlow framework using GeForce RTX 3080 GPU. We evaluated OCEAN-AI on the ChaLearn First Impressions v2 (FI v2) corpus [1] and found that it outperformed all other known methods [7] based on processing these three modalities. Using our framework, we also obtained initial baselines for our MuPTA corpus [3]. Model training details are described in [2, 3].

he/she uploads files and clicks the button to calculate PT (see Figure 2, top). Next, the user selects a practice task and clicks the button to complete it (see Figure 2, bottom part). The user then receives a ranked information displaying the most suitable person according to the criteria set. Our app operates in real-time, taking 0.6 sec to process a 1 sec video using Intel i9 CPU.

3. Conclusions

Automatic PTA is crucial for various aspects of life, especially for tasks of recruiting, team building, product customization and so on. Our OCEAN-AI framework integrates three modules for extracting heterogeneous video, audio, and text features of human’s behavior. The multimodal feature fusion module applies the Siamese network with the gated attention mechanism. The implementation of solutions for ranking candidates and building efficient teams in OCEAN-AI demonstrates its high practical value. Potential applications of OCEAN-AI also include an analysis of personality disorders, personalized intelligent assistants and individual educational trajectories.

4. Acknowledgements

This work was supported by the Analytical Center for the Government of the Russian Federation (IGK 000000D730324P540002), agreement No. 70-2021-00141.

5. References

- [1] H. J. Escalante *et al.*, “Modeling, recognizing, and explaining apparent personality from videos,” *IEEE Transactions on Affective Computing*, vol. 13, no. 2, pp. 894–911, 2020.
- [2] E. Ryumina *et al.*, “OCEAN-AI framework with emofomer cross-hemiface attention approach for personality traits assessment,” *Expert Systems with Applications*, vol. 239, p. 122441, 2024.
- [3] —, “Multimodal personality traits assessment (MuPTA) corpus: The impact of spontaneous and read speech,” in *Proceedings of INTERSPEECH*, 2023, pp. 4049–4053.
- [4] P. D. Tieger *et al.*, *Do what you are: Discover the perfect career for you through the secrets of personality type*. Hachette UK, 2014.
- [5] A. Furnham, “The big five facets and the MBTI: The relationship between the 30 neo-pi (r) facets and the four myers-briggs type indicator (MBTI) scores,” *Psychology*, vol. 13, no. 10, pp. 1504–1516, 2022.
- [6] E. Ryumina *et al.*, “In search of a robust facial expressions recognition model: A large-scale visual cross-corpus study,” *Neurocomputing*, vol. 514, pp. 435–450, 2022.
- [7] D. Giritlioglu *et al.*, “Multimodal analysis of personality traits on videos of self-presentation and induced behavior,” *Journal on Multimodal User Interfaces*, vol. 15, no. 4, pp. 337–358, 2021.

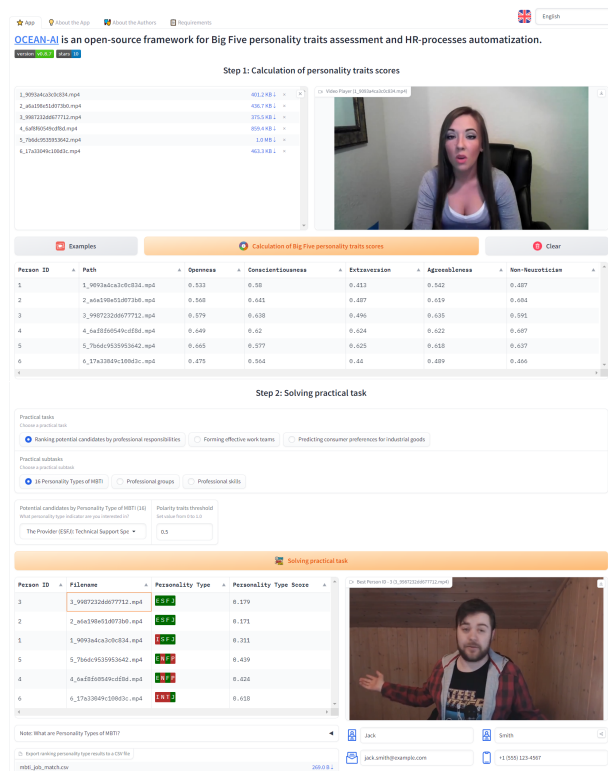


Figure 2: A Multimodal user interface of OCEAN-AI. Examples are taken from the publicly available FI v2 [1] corpus.

Source codes⁷ and a software documentation⁸ of OCEAN-AI are free-available. A multimodal user interface (hereafter an app) was implemented using Gradio⁹ on the Hugging Face platform. To use the app, a user should perform two steps. Firstly,

⁷<https://github.com/aimclub/OCEANAI>

⁸<https://oceanai.readthedocs.io/en>

⁹<https://huggingface.co/spaces/ElenaRyumina/OCEANAI>