APPARENT PERSONALITY ANALYSIS BASED ON AGGREGATION MODEL

Milić Vukojičić*,
Mladen Veinović

Singidunum University,
Belgrade, Serbia

Abstract:

Apparent traits personality analysis based on multimodal traits from text, handwriting, images, and audio is a challenging problem in computer vision, signal processing, and deep learning. To improve models based just on one of the input parameters we will combine all four of them with the aggregation layer. Models based on handwriting and images are the best predictors of the Neo-PI-R results/profile derived from the results of the NEO-PI-R. In addition, to get the best results we showed different aggregation layers (Max, Min, Median, Mean). We obtain the highest prediction certainty for Consciousness, Extraversion, Agreeableness, and Neuroticism. While Openness to experience was very hard to predict with the use of the aggregation model. As the five traits are not homogenous, but consist of facets that do not necessarily converge, and deeper analysis of the facets shows that the score on the main trait is nothing more than the mean of the facets scores, and that limitation could be overcome by analyzing the facets' behavior and predictability. This limitation can be overcome with further research done in the domain of apparent personality analysis with traits and their facets.

Keywords:

Apparent personality analysis, personality classification, feature classification, aggregation functions.

INTRODUCTION

The first impression based on Big Five Personality Traits, or so-called OCEAN (Openness to Experience, Consciousness, Extraversion, Agreeableness, and Neuroticism) is one of the most frequently and explored models in human trait analysis. Vinciarelli & Mohammadi [1] showed in their paper that the automatic recognition, perception, and synthases of personality are some of the hardest domains of Computer Science. The personality traits approach is one of the most promising ones to assess personality characteristics/features, and the BIG-5 model is one of the most explored and possibly promising ones. The best model in the field of trait prediction is the model based on multiple input parameters, called multimodal trait prediction.

Correspondence:
Milić Vukojičić

e-mail:
vukojicic.milic@gmail.com

DOI: 10.15308/Sinteza-2021-220-225

This model is trying to have many important aspects of a person as an input, such as a person's handwriting, a person's digital text, a still image of a person or image sequence, and an audio signal of the person's voice. Gavrilescu & Vizireanu [2] showed that state-of-the-art models based on handwriting have an accuracy of more than 80% for all traits. Ponce-López & Chen [3] showed that some traits are easier to predict than others, such as "Agreeableness" which is very hard to predict out of the video material. State-of-the-art models have more than 90% accuracies as shown in the paper written by Zhang C. L. and Zhang H. [4].

They propose Deep Bimodal Regression (DBR) framework. Their model is based on the video which is broken down to images and audio features, where they combine the scores of the visual modality and scores of the audio modality. Studies in psychology are dealing with behavior (B), which is represented as a function of the person (P) at the given situation (S) [5], and apparent personality (A) is conditioned by the observer (O) and that is described by the function A = f(P, S, O). Our approach is trying to minimize the subjectiveness of the observer [6], and also the results shown by Vinciarelli [7] are allowing us to create statistical models of automatic personality computing. The main assumption is that traits are not rigid and if we want to have a better and more realistic model we need to combine more than two models. The main limitation of the previous research is that they are based on a small amount of the input parameters. In this paper, we are trying to combine more than two inputs and create the multimodal model with the combination of text, handwriting, image sequence, and audio-visual traits and a suitable aggregation function, aggregation layer which goal are to predict OCEAN profile better than a single input model or models based on the single neural network. The purpose of the aggregation layer is to convert an array with spatial dimensions to a vector that has a fixed size.

Aggregation functions that are represented in this paper are Mean, Median, Max, and Min. We have shown that the aggregation function can give us a better approximation of traits and give us more accurate results when we compare it to NEO Personality Inventory (NEO PI-R) test.

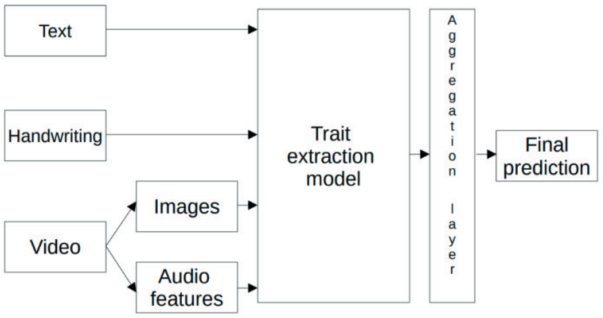

Figure 1. Framework of the proposed model with Aggregation layer, video is treated as having two natural modalities, and text and handwriting one natural modality, the final predicted personality traits are created from the aggregation layer by the fusion of all outputs of text, handwriting, images and audio features.

The challenging part of the personality analysis is connected to many factors that can affect results such as individual and cultural differences, observation conditions, random noise that can be found in camera and microphone, different styles of articulation, etc.

The proposed solution of combining four modalities with an aggregation layer strives to give a better result and to minimize factors that can affect the result. The aggregation layer will be created out of the aggregation function that we can find in the work of Grabisch & Marichal [8].

2. RELATED WORK

In this section, we will review the work that is related to apparent personality analysis from text, handwriting, images, and audio feature using deep learning.

2.1. PERSONALITY DETECTION FROM TEXT WITH MACHINE LEARNING

Deep learning methods that are used in previous papers are based on seven layers: input layer used for word vectorization and convolution used for sentence vectorization combined with max-pooling layer. After that, we have 1-max pooling for document vectorization and concatenation layer and linear with Sigmoid activation for classification and two-neuron softmax output layers [9,10]. The best approach as an input parameter to textual modality will be to uses vectorized words from wod2vec models [11,12,13].

Models proposed only from the text for extracting traits have lower results than other models. Best results are shown in previous works when CNN is used with document features. The same work is showing us that we cannot get better results if we use n-grams.

The best results are showed with the combination of Mairesse[14] and CNN[9] with accuracy O = 62.68, C = 57.30, E = 58.09, A = 56.71 and N = 59.38.

The advantage of this solution is that we need little data about the subject to predict their traits, which is why this model is very popular among the papers that are based on predicting traits on social network platforms. The drawback of a solution based on text is very close to 50% or very close to random prediction. To be useful this model needs to be combined with other models which are more precise in their prediction. The output of the majority of models proposed in several papers is binarised [0,1] which leads to a low prediction score and with a combination of accuracy of prediction it can lead to false predictions.

2.2. PERSONALITY DETECTION FROM HANDWRITING WITH MACHINE LEARNING

Handwriting analysis is a scientific method of understanding personality traits from patterns present in handwriting also from strokes. Paper proposed by Champa & AnandaKumar based on a set of rule-based classifiers has an accuracy of prediction rate of 68% [15].

Non-intensive three-layer architecture [2] based on offline handwriting offers prediction accuracies of 84.40% for Openness to Experience, Extraversion, and Neuroticism to 70% for Conscientiousness and Agreeableness. Very close to this model are Support Vector-Machines, k-nearest neighbors, and AdaBoost classifiers with 72% accuracy proposed from Chen & Lin [16] and Hidden Markov Models and neural networks Fallah and Khotanlou [17] with 78% accuracy.

Predicting personality, OCEAN traits from handwriting is the oldest scientific method. A drawback of this method is that we need to have a written message from the subject.

2.3. VISUAL-BASED PERSONALITY EXTRACTION WITH MACHINE LEARNING

Visual-based personality extraction is mainly done with the part of machine learning called deep learning and reinforcement learning. Deep learning is mainly related to the work in the domain of Convolutional Neural Networks [18,19] wherefrom each image or image sequence we can extract different traits. A different approach than this is mainly done in the domain of reinforcement learning and the long-short term memory approach [20]. Deep Bimodal Regression is represented in a method based on deep regression for the visual modality [4].

	Accuracy of prediction
NJU-LAMNBDA[4]	91.1
LSTM[18]	69.4
GME-LSTM[18]	76.5

Table 1. Comparison of visual-based personality extraction with convolutional neural networks, reinforcement learning, and long-short term memory and Deep Bimodal Regression.

The advantage of this solution is that prediction is very accurate, and most of the time with state-of-the-art approaches. A drawback is that we need many images for training, and the classification of the images can be very hard. Therefore, this is why databases for this specific problem are hard to find. Some of the predictions within the traits like Openness to experience and Neuroticism are hard to detect accurately.

2.4. AUDIO-BASED PERSONALITY EXTRACTION WITH MACHINE LEARNING

The main methods of predicting audio-based features are time domain and frequency domain features. Linear Prediction Cepstral Coefficient (LPCC) [21] and Mel Frequency Cepstral Coefficients (MFCC) [20] are based on short-term spectral-based sound features which are obtained from spectrum-of-a-spectrum of audio files [4].

Deep Bimodal Regression uses a log filter bank called log bank features based on Mel Frequency Cepstral Coefficients (MFCC) and results represented with the DBR are outperforming MFCC results by 0.75% with an accuracy of prediction of 89%. In the majority of papers, this method is used if we want to find binarised output [0,1], and this will lead to unrealistic predictions.

3. THE PROPOSED AGGREGATION METHOD

Aggregation method that we want to explore in this paper is based on several Aggregation functions. Aggregation function is the function that is the process of merging multiple values into a single value which in this case will represent our trait. The main problem with previous methods is that all of the methods are good at predicting some of the traits while some traits are not represented in the way that we want. For example, it is very easy to predict Neuroticism from handwriting and very hard to predict this trait from images [2,4]. Openness as a trait is very hard to predict from handwriting and images and videos are giving us better results [2,4,15,16,17,19,20]. The output value of the aggregation function needs to compute a value that will represent all possible inputs and give us a very close, sometimes near-perfect approximation of the real input. The input of the aggregation layer can be represented as a matrix of size 5x4, and this will be the same for all aggregation layers, and with the output represented as a vector of size 5. The matrix that we have on input will be represented with values 0≤*input*≤1 (1) and the vector that we have as output will be represented with values 0≤*output*≤1 (2).

$$M_{input} = \begin{bmatrix} T_o & H_o & I_o & A_o \\ T_c & H_c & I_c & A_c \\ T_a & H_a & I_a & A_a \\ T_n & H_n & I_n & A_n \end{bmatrix} \tag{1}$$

$$V_{output} = [O\,C\,E\,A\,N] \tag{2}$$

Matrix will be represented o through the values extracted from the Trait extraction model shows in Figure1. Variables $\{T_o, T_c, T_e, T_a, T_n\}$ will represent personal traits detected from text, $\{H_o, H_c, H_e, H_a, H_n\}$ are OCEAN traits detected from handwriting. Traits from images will be represented as $\{I_o, I_c, I_e, I_a, I_n\}$ and $\{A_o, A_c, A_e, A_a, A_n\}$ are the set of audio-based traits.

Vector *Voutput* will represent the final prediction of all traits O - Openness to experience, C - Consciousness, E - Extraversion, A - Agreeableness, and N – Neuroticism, and output values will be represented in the range of 0≤*output*≤1.

Min and Max aggregation functions are represented with the formula (3) (4). The output of the min function should yield the numerically smallest of the numbers between $\{x_1,\dots,x_n\}$. The output of the max function should yield the numerically largest of the numbers between $\{x_1,\dots,x_n\}$. In literature, we can find that minimum and maximum functions can be represented using the lattice operations ∧,∨ shown in formula (3) (4)

$$Min(x) = \wedge_{j=1}^{n} x_i \tag{3}$$

$$Max(x) = \vee_{j=1}^{n} x_i \tag{4}$$

The next function that is represented will be the arithmetic mean function AM: $I^n \rightarrow I$ (5). The mean function is also known as the expectation function or average function.

$$AM(x) := \frac{1}{n}\sum_{i=1}^{n} x_i \tag{5}$$

The median aggregation layer should return the median of the elements in a list. To get the median of the list we need first to have a sorted list, or the average of the two center elements if the list is of even length. We need to make a difference if the input to our median function is an odd number of elements (6) or is an even number of elements (7).

$$OddMed(x_1,\dots,x_{2k1}) := x_{(k)} = \bigwedge_{\substack{K\subseteq[2k-1] \\ |K|=k}} \bigvee_{i\in K} x_i = \bigvee_{\substack{K\subseteq[2k-1] \\ |K|=k}} \bigwedge_{i\in K} x_i \tag{6}$$

$$EvenMed(x_1,\dots,x_{2k1}) := AM(x_{(k)}, x_{(k+1)}) = \frac{x_{(k)} + x_{(k+1)}}{2} \tag{7}$$

4. EXPERIMENTS

The only problem is the lack of a database that contains text, handwritings, and video and extracted traits for the model proposed in Figure 1. To support the study, we build our database. The database contains a collection of text, handwriting, and video, including images and audio features from 64 subjects (31 females and 33 males), where the subject age is between 18 and 70 years old, as well as their result of the NEO PI-R test from which personality traits are evaluated correctly. All subjects took the NEO-PI-R test and the main evaluation of our model of prediction will be when we compare results of the NEO-PI-R test with the results of the model proposed in Figure1. Also, we collect a short video sequence with audio from all subjects and an example of their handwriting as well as a short essay.

For the text, we used a model based on CNN + Mairesse [9] and for handwriting, we used non-intensive three-layer architecture [2]. The text is based on vectorized words from wod2vec models [10,11,12,13], and audio–based features are based on Mel Frequency Cepstral Coefficients (MFCC) [20], with values represented with 1 or 0. We need to be very careful how we are going to use text and audio-based traits in the final prediction because of the polarity between 0 and 1.

Traits based on	Absolute error[Δx]				
	O	C	E	A	N
Image	0.03	**0.01**	0.03	0.03	**0.01**
Audio	0.14	0.11	0.10	0.03	0.06
Handwriting	**0.02**	**0.01**	0.02	0.02	0.03
Text	0.24	0.17	0.03	0.07	0.10
Max aggregation	0.27	0.29	0.35	0.32	0.34
Min aggregation	0.55	0.46	0.46	0.45	0.48
Mean aggregation	0.11	0.07	**0.01**	0.03	0.04
Median aggregation	0.07	0.05	0.04	**0.01**	0.02

Table 2. Results of proposed aggregation model compared with NEO-PI-R results and different trait prediction models, absolute error.

Traits based on	Relative error[δx]				
	O	C	E	A	N
Image	2%	9%	14%	3%	5%
Audio	9%	13%	10%	4%	2%
Handwriting	**1%**	4%	8%	3%	5%
Text	31%	25%	15%	10%	8%
Max aggregation	52%	60%	72%	63%	68%
Min aggregation	81%	79%	75%	71%	77%
Mean aggregation	10%	6%	**7%**	**1%**	**1%**
Median aggregation	5%	**3%**	15%	5%	5%

Table 3. Results of proposed aggregation model compared with NEO-PI-R results and different trait prediction models, relative error.

Results showed in Table 2 and Table 3 are the representation of absolute and relative error based on the result from the NEO-PI-R test and different models. Models based on handwriting and models based on images are models which are closest to the result of NEO-PI-R tests. Proposed models based on Max and Min aggregation are models that have the highest error of prediction when we compare them to the NEO-PI-R model. Models based on Mean and Median have an improvement over models based on handwriting and images. Openness to Experience is best predicted from models based on handwriting [Δx]=0.01,[δx]=1"\%" , while the models proposed in this paper have a large deviation of NEO-PI-R results.

Model-based on Mean aggregation gives the best prediction when it comes to Consciousness[Δ*x*]=0.08,*[δx*]=6"\%". Extraversion [Δ*x*]=0.03,[*δx*]=7"\%", Agreeableness[Δ*x*]=0.03,[*δx*]=1"\%", and Neuroticism[Δ*x*]=0.04,[*δx*]=1"\%" are best predicted from the model based on Median aggregation.

5. CONCLUSION

Apparent personality analysis based on multimodal traits is an extensive problem in computer vision and machine learning research. To improve results based on a different state-of-the-art model for apparent personality prediction, this paper has proposed different aggregation functions or layers built in the multimodal trait prediction. The revised NEO-PI-R [22] is one of the best instruments used for the prediction of Openness to Experience, Consciousness, Extraversion, Agreeableness, and Neuroticism. The solution showed in this paper showed that we can get very close predictions from multimodal models based on Mean and Median aggregation functions, for every trait apart from Openness to experience.

Feature work in the domain of apparent personality analysis can be improved in the domain of prediction not only BIG-5 traits but also their facets [23]. This can lead to a more accurate model and a better representation of the subject's traits.

REFERENCES

[1] A. Vinciarelli and G. Mohammadi, "A survey of personality computing," IEEE Transactions on Affective Computing, vol. 5, no. 3, pp. 273-291, 2014.

[2] M. Gavrilescu and N. Vizireanu, "Predicting the Big Five personality traits from handwriting," EURASIP Journal on Image and Video Processing, vol. 2018, no. 1, pp. 1-17, 2018.

[3] V. Ponce-López et al., "Chalearn lap 2016: First round challenge on first impressions-dataset and results," in European conference on computer vision, 2016: Springer, pp. 400-418.

[4] C.-L. Zhang, H. Zhang, X.-S. Wei, and J. Wu, "Deep bimodal regression for apparent personality analysis," in European conference on computer vision, 2016: Springer, pp. 311-324.

[5] J. C. S. J. Junior et al., "First impressions: A survey on vision-based apparent personality trait analysis," IEEE Transactions on Affective Computing, 2019.

[6] D. C. Funder, "Accurate personality judgment," Current Directions in Psychological Science, vol. 21, no. 3, pp. 177-182, 2012.

[7] A. Vinciarelli, "Social perception in machines: The case of personality and the Big-Five traits," in Toward Robotic Socially Believable Behaving Systems-Volume II: Springer, 2016, pp. 151-164.

[8] M. Grabisch, J.-L. Marichal, R. Mesiar, and E. Pap, Aggregation functions (no. 127). Cambridge University Press, 2009.

[9] N. Majumder, S. Poria, A. Gelbukh, and E. Cambria, "Deep learning-based document modeling for personality detection from text," IEEE Intelligent Systems, vol. 32, no. 2, pp. 74-79, 2017.

[10] S. Park, H. S. Shim, M. Chatterjee, K. Sagae, and L.-P. Morency, "Computational analysis of persuasiveness in social multimedia: A novel dataset and multimodal prediction approach," in Proceedings of the 16th International Conference on Multimodal Interaction, 2014, pp. 50-57.

[11] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," arXiv preprint arXiv:1301.3781, 2013.

[12] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," arXiv preprint arXiv:1310.4546, 2013.

[13] T. Mikolov, W.-t. Yih, and G. Zweig, "Linguistic regularities in continuous space word representations," in Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: Human language technologies, 2013, pp. 746-751.

[14] S. M. Mohammad and S. Kiritchenko, "Using hashtags to capture fine emotion categories from tweets," Computational Intelligence, vol. 31, no. 2, pp. 301-326, 2015.

[15] H. Champa and K. AnandaKumar, "Artificial neural network for human behavior prediction through handwriting analysis," International Journal of Computer Applications, vol. 2, no. 2, pp. 36-41, 2010.

[16] Z. Chen and T. Lin, "Automatic personality identification using writing behaviours: an exploratory study," Behaviour & Information Technology, vol. 36, no. 8, pp. 839-845, 2017.

[17] B. Fallah and H. Khotanlou, "Identify human personality parameters based on handwriting using neural network," in 2016 Artificial Intelligence and Robotics (IRANOPEN), 2016: IEEE, pp. 120-126.

[18] M. Chen, S. Wang, P. P. Liang, T. Baltrušaitis, A. Zadeh, and L.-P. Morency, "Multimodal sentiment analysis with word-level fusion and reinforcement learning," in Proceedings of the 19th ACM International Conference on Multimodal Interaction, 2017, pp. 163-171.

[19] S. Poria, H. Peng, A. Hussain, N. Howard, and E. Cambria, "Ensemble application of convolutional neural networks and multiple kernel learning for multimodal sentiment analysis," Neurocomputing, vol. 261, pp. 217-230, 2017.

[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," Advances in neural information processing systems, vol. 25, pp. 1097-1105, 2012.

[21] S. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," IEEE transactions on acoustics, speech, and signal processing, vol. 28, no. 4, pp. 357-366, 1980.

[22] P. T. Costa Jr and R. R. McCrae, The Revised NEO Personality Inventory (NEO-PI-R). Sage Publications, Inc, 2008.

[23] C. G. DeYoung, L. C. Quilty, and J. B. Peterson, "Between facets and domains: 10 aspects of the Big Five," Journal of personality and social psychology, vol. 93, no. 5, p. 880, 2007.