Face-Based Automatic Personality Perception

Noura Al Moubayed, Yolanda Vazquez-Alvarez, Alex McKay, Alessandro Vinciarelli
University of Glasgow (School of Computing Science) - Sir A.Williams Building, G128QQ Glasgow (UK)
firstname.lastname@glasgow.ac.uk

ABSTRACT

Automatic Personality Perception is the task of automatically predicting the personality traits people attribute to others. This work presents experiments where such a task is performed by mapping facial appearance into the Big-Five personality traits, namely Openness, Conscientiousness, Extraversion, Agreeableness and Neuroticism. The experiments are performed over the pictures of the FERET corpus, originally collected for biometrics purposes, for a total of 829 individuals. The results show that it is possible to automatically predict whether a person is perceived to be above or below median with an accuracy close to 70 percent (depending on the trait).

Categories and Subject Descriptors: J.4 [Computer Applications]: social and behavioral sciences; H.4.3 [Information Systems Applications]: Communications Applications. **General Terms:** Human Factors, Experimentation. **Keywords:** Automatic Personality Perception, Eigenfaces, Big-Five.

1. INTRODUCTION

The social cognition literature shows that people attribute personality traits to others in less than one second after the first contact [11]. The main interest of attributed traits is that they are predicitve of how others behave towards a certain person and, furthermore, they explain better than self-assessed traits the social characteristics of an individual [14]. Trait attribution was shown to take place not only during face-to-face encounters, but also during the consumption of multimedia material (e.g. when people see someone on television), an effect known as *Media Equation* [6]. Therefore, a large body of previous research work has aimed at automatically predicting the traits people attribute to unacquainted others they observe in multimedia material [12].

So far, research on Automatic Personality Perception (APP) has focused mainly on nonverbal behavior (paralanguage, fidgeting, gaze, etc.) and online activities (tagging images

as "*favorite*" on Flickr, Facebook profiles, etc.) [12]. In parallel, research on face analysis has targeted the recognition of facial expressions [2], emotions and the inference of attractiveness [3]. The problem of automatically mapping facial appearance into personality traits has been largely neglected in the computing literature (see [12] for the few works that consider the problem). This work aims at filling, at least partially, such a gap and proposes experiments aimed at predicting whether people perceive an individual portrayed in a picture to be above or below median with respect to each of the Big-Five Traits [5].

The experiments have been performed over the frontal face images of the Color FERET database, a corpus originally collected for biometrics purposes (see Section 2). The most challenging aspect of such a corpus is the wide variety in terms of ethnicity, clothes, haircut and any other characteristics that contribute to facial appearance. From this point of view, this work is an advancement with respect to previous experiments that have focused on synthetic faces [7]. In particular, the experiments of this work can be considered more realistic and representative of everyday situations where people observe the portraits of unknwon individuals (e.g., profile pictures on social media, testimonials in advertisements, etc.).

The experiments involve a large sample of 829 individual faces. The largest corpus for face-based APP presented so far in the literature includes only 300 different faces generated artificially [7]. The main advantage in increasing the number of facial images is not only that the statistical reliability of the results improves, but also that it is possible to observe weaker effects (e.g., correlation between features and traits) that might not reach statistical significance with a smaller number of facial images.

The rest of the paper is organized as follows: Section 2 describes the data collection process, Section 3 presents the proposed approach, Section 4 reports on experiments and results, and the final Section 5 draws some conclusions.

2. THE DATA: FACE AND PERSONALITY

The corpus used for the experiments is a selection of 829 images of real faces (a total of 829 different individuals) from the Color FERET database[1], a benchmark for face recognition technologies collected between December 1993 and August 1996. All the selected images are in color and portray individuals of different gender (478 males and 351 females) and ethnic background (63.9% White, 15.3% Asian/Pacific

---

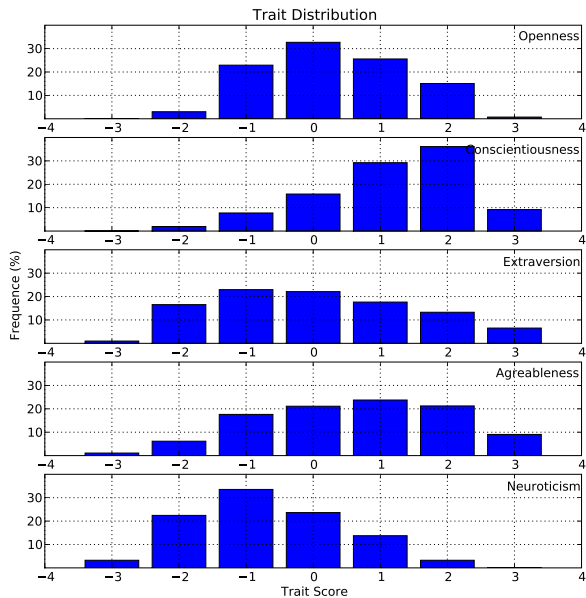[1] http://www.nist.gov/itl/iad/ig/colorferet.cfm

Figure 1: Score distribution for the Big-Five traits.

Islands, 7.6% African, 7.4% Indo-Aryan, 4.6% Hispanic/Latino, 1.1% Arab and 0.1% Native American). The personality assessments were provided by 11 independent judges (5 females, 6 males, aged 20 to 39 years) of White British ethnic background, with normal vision and no hearing problems. They were all paid for their participation. The reason behind the use of several assessors is that the goal of APP is to predict what is common across the traits attributed by different observers [12] and not to predict how one person in particular perceives others. On the other hand, all assessors have the same cultural background in order to avoid noise due to cultural effects.

Each judge filled in the *Big Five Inventory 10* (BFI-10) [5] questionnaire for all 829 images in the database. The BFI-10 includes 10 items (e.g., "*I see this person as someone who is lazy*", "*I see this person as someone who tends to find faults with others*", etc.) and every item of the questionnaire can be rated on a scale from 1:5 (ranging from "Disagree Strongly" to "Agree Strongly"). By summing the answers given to the various items, it is possible to obtain five scores that account for how high a facial image is perceived to be along each of the Big-Five personality dimensions, namely *Openness* (tendency to be curious, creative, artistic, etc.), *Conscientiousness* (tendency to be trustworthy, reliable, efficient, etc.), *Extraversion* (tendency to be sociable, active, etc.), *Agreeableness* (tendency to be kind, generous, etc.) and *Neuroticism* (tendency to be anxious, tense, etc.). Figure 1 shows the observed distribution of scores for the various traits.

The assessments were collected through an online system allowing the judges to see sequentially the 829 pictures and, for each of them, to fill the BFI-10 questionnaire. A judge was allowed to move from one picture to the next only when all BFI-10 items had been answered. The face images were presented in a different randomised order to each judge. After an initial face-to-face introductory session with the experimenters, the judges were asked to fill out the online

questionnaires in no more than four weeks. This allowed the judges to complete the assessments in separate sessions (they were asked to work no more than 30 minutes per session in order to avoid tiredness effects), in their own time, without any contact with the other judges and/or the experimenters.

At the end of the assessment collection, there were 11 filled questionnaires per picture (one per judge). The scores resulting from each of them were averaged to provide the final personality assessment. The consistency between the judges was measured in terms of *effective reliability* ρ [8]:

$$\rho = \frac{Nr}{1 + (N-1)r}, \qquad (1)$$

where r is the average inter-judge correlation and N is the total number of judges. The ρ value was 0.80 for Openness, 0.86 for Conscientiousness, 0.90 for Extraversion, 0.89 for Agreeableness and 0.79 for Neuroticism. The values are sufficient to achieve accuracies above chance level in the APP task presented in this work (see Section 4).

3. THE APPROACH

The approach proposed in this work includes three main steps, namely face preprocessing, feature extraction and trait prediction. During the first step, the images are converted into grey scale and undergo histogram equalization to normalize the brightness and increase the contrast. The Viola-Jones detector [13] is applied to the resulting images to locate a rectangular region around the face. The rectangle is then extended by 20% on the bottom and 10% on both left and right sides to include neck and ears. The extended rectangles are extracted from the original images and scaled to 65×50 pixels. The goal is to ensure that all samples have the same shape and size while reducing the computational cost of the following steps.

This approach adopts appearance features (see below) that are sensitive to variations in the positions of salient points such as eyes, nostrils, chin tip, lip corners, etc. For this reason, the samples are mapped onto a reference image that corresponds to the average of the corpus. This task is performed using the non-rigid registration approach[2] proposed in [9].

The second step adopts the eigenfaces method [10], one of the face representation techniques most commonly adopted in literature. The eigenfaces are the Principal Components extracted from a set of training images and they constitute a basis in a space where every point corresponds to a different face image. The eigenfaces are ordered according to the amount of variance they account for in the original data. In the experiments presented in this paper, the first 103 eigenfaces correspond to 90% of the variance. A test face is represented with its projections onto the eigenfaces and the exact number of eigenfaces to be retained is set through crossvalidation (see Section 4).

The trait prediction step is performed using Support Vector Machines (SVMs) with both polynomial and Gaussian Radial Basis Function (RBF) kernel [1]. The SVMs were trained with the Sequential Minimal Optimization approach[3]

---

[2]Implementation available at `http://www.sas.el.utwente.nl/open/research/medical_imaging/registration`

[3]Implementation available at `http://www.mathworks.co.uk/help/stats/support-vector-machines-svm.html`

and the hyperparameters were set through crossvalidation (see Section 4).

4. EXPERIMENTS AND RESULTS

The experiments were performed over 785 images because the face detection approach did not work correctly for 44 samples (5.5% of the total). The tests followed a K-fold experimental protocol ($K = 20$). The data was first split into K subsets of roughly the same size, then $K - 1$ subsets were used for training while the remaining one was left-out for testing. The process was iterated K times and, at each repetition, a different subset was used for testing. The main advantage of this protocol is that it allows one to use all the data at disposition to test an approach while still ensuring a rigorous separation between training and test set.

The hyperparameters were set through crossvalidation, the values leading to the highest accuracy over the training set were used for the test. The hyperparameters include the number d of eigenfaces to be retained, the order n of the polynomial kernel and the scaling factor γ for the RBF kernel. The values of d, n and γ were searched in the range $[10, 100]$, $[1, 5]$ and $[1, 10^5]$, respectively.

Figure 2 shows the Spearman correlation coefficients between the Big-Five Traits and the projections of the face images onto the first 10 eigenfaces. These latter account for 60% of the data variance and appear at the bottom of the figure. On average, the performance of automatic prediction approaches tends to be higher when the features (or at least a sufficient number of them) covariate to a statistically significant extent with the traits. Furthermore, the correlational analysis can provide indications on the features most likely to influence the outcome of the prediction approaches.

Overall, the Spearman coefficients tend to be weak (the maximum value was observed for Openness and it is 0.3). However, the p-value is lower than 5% for 22% of the projections for Openness, 10% for Conscientiousness, 28% for Extraversion, 26% for Agreeableness and 15% for Neuroticism. This seems to suggest that the covariation between projections and traits is not just the effect of chance (in this case no more than 5% of the features would have p-value lower than 5%), but the result of a consistent relationship between face appearance and personality traits.

The experiments presented in this paper address two tasks: the first is to predict whether an individual is perceived to be *high* (above or equal to the median) or *low* (below median) along each of the Big-Five Traits, the second is to predict whether an individual is in the *top* or *bottom* quartile for each of the Big-Five Traits. The first task is performed using the entire corpus of images at disposition, the second using only half of the data (the images falling in the central quartiles of each trait are discarded). In both cases, the performance is measured in terms of *accuracy*, i.e. percentage of times the approach assigns a face image to the right class.

The results for the first task are reported in the upper plot of Figure 3 using both polynomial and Radial Basis Function (RBF) kernels for the Support Vector Machines. The error bars correspond to the standard deviation of the accuracy across the K folds. According to a two-tailed binomial test, the accuracy difference is statistically significant ($p < 0.01$) with respect to chance (a baseline system that predicts always the class with the highest a-priori probability). The only exception is the performance of the SVMs with poly-
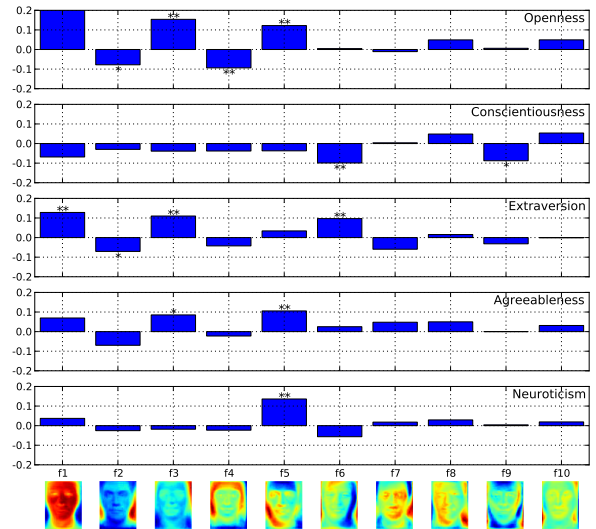

Figure 2: The charts show the Spearman correlation coefficients between the projections onto the first 10 eigenfaces (at the bottom of the figure) and the Big-Five Traits. Overall, the first 10 eigenfaces account for 60% of the variance. The asterisks indicate whether the coefficient is statistically significant at 5% (single asterisk) or 1% (double asterisk) confidence level.

nomial kernel in the case of Conscientiousness. The RBF kernel systematically increases the performance with respect to the polynomial one, but not to a statistically significant extent (according to a two-tailed t-test). The best results are achieved for Openness, Neuroticism and Extraversion. This latter tends to be the best predicted trait in most of the APP experiments presented in the literature [12]. The probable reason is that such trait has been shown to be the one people tend to attribute more easily in zero acquaintance scenarios like the one considered in this work [4]. In this respect, the results seem to confirm the indications of personality psychology.

Experiments similar to those presented in this paper were proposed in [7] over synthetic faces. However, the experiments of such work take into account only samples for which the traits are in the top and bottom quartile of the observed personality scores. Therefore, for the sake of comparison, the first experiment above was repeated with this setup. The main reason for retaining only the judgments at the extremes of the distribution is that these result from higher consensus among the assessors. Hence, the relationship between facial appearance and attributed traits should be more consistent for the samples in the top and bottom quartiles.

The results are reported in the lower plot of Figure 3. The accuracy increases for all traits and for both types of kernel adopted in the SVM. However, the improvement does not appear to be statistically significant (according to a two-tailed t-test). This seems to suggest that the scores belonging to the central quartiles do not result from the averaging over different scores (e.g., the rating 3 can be obtained be-
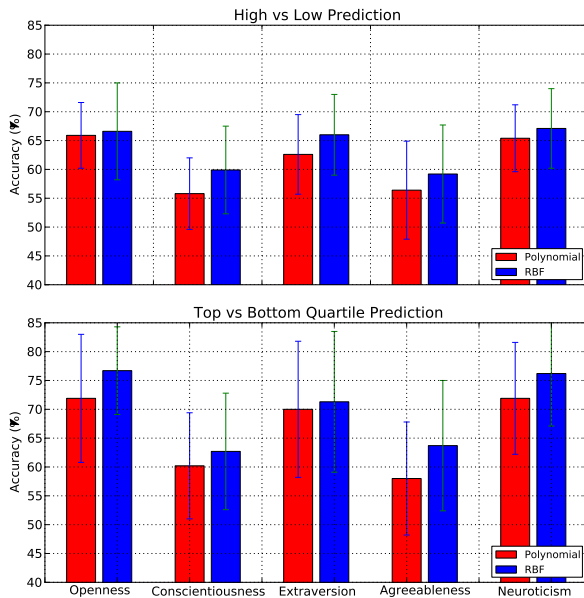
Figure 3: The chart shows the accuracies achieved using Support Vector Machines with polynomial and RBF kernel.

cause all assessors assign 3, but also because some assign 2 and others assign 4), but from substantial agreement across the judges. Such a hypothesis is further corroborated by the reliability values measured in Section 2.

A rigorous comparison between the results of this work and those presented in [7] is not possible because data and trait set are different. However, some orientative indications can still be obtained by taking into account three traits adopted in [7] that can be reasonably matched with one of the Big-Five traits, namely *Extroverted* (corresponding to Extraversion), *Competent* (corresponding to Conscientiousness) and *Likable* (corresponding to Agreeableness). For these three traits, the accuracies are similar to those obtained in this work (when using the same approach as the one adopted here). While not being fully rigorous, the comparison still seems to suggest that the trait attribution process does not change significantly when passing from synthetic to natural images.

5. CONCLUSIONS

This paper has presented experiments on face-based Automatic Personality Perception, i.e. the prediction of the traits people attribute to unacquainted others based on facial appearance. To the best of our knowledge, this is one of the first works dedicated to this problem, previous research on faces having targeted mainly facial expressions [2], emotions and attractiveness [3]. The results show that it is possible to predict with accuracy of around 65% whether an individual portrayed in a picture is perceived to be *high* (above or equal to the median) or *low* (below the median) along Openness, Extraversion and Neuroticism. The accuracy is lower, but significantly above chance, for the other two traits (Conscientiousness and Agreeableness).

Future work will adopt face analysis approaches more sophisticated than the one adopted here, especially when it

comes to structural properties such as the distribution of salient facial points (eyes, nostrils, lip corners, etc.). Moreover, the corpus collected for this work gives the possibility of exploring the effect of clothes, hair color, ethnic background and other factors that so far have been largely neglected in the literature.

6. REFERENCES

[1] C. Cortes and V. Vapnik. Support vector machine. *Machine learning*, 20(3):273–297, 1995.

[2] B. Fasel and J. Luettin. Automatic facial expression analysis: a survey. *Pattern Recognition*, 36(1):259–275, 2003.

[3] H. Gunes. A survey of perception and computation of human beauty. In *Proceedings of the Workshop on Human Gesture and Behavior Understanding*, pages 19–24, 2011.

[4] C. Judd, L. James-Hawkins, V. Yzerbyt, and Y. Kashima. Fundamental dimensions of social judgment: Unrdestanding the relations bteween judgments of competence and warmth. *Journal of Personality and Social Psychology*, 89(6):899–913, 2005.

[5] B. Rammstedt and O. John. Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality*, 41(1):203–212, 2007.

[6] B. Reeves and C. Nass. *The media equation*. Cambridge University Press, 1996.

[7] M. Rojas, D. Masip, A. Todorov, and J. Vitria. Automatic prediction of facial trait judgments: Appearance vs. structural models. *PLoS One*, 6(8):e23323, 2011.

[8] R. Rosenthal. Conducting judgment studies: Some methodological issues. In J. Harrigan, R. Rosenthal, and K. Scherer, editors, *The new handbook of methods in nonverbal behavior research*, pages 199–234. Oxford University Press, 2005.

[9] D. Rueckert, P. Aljabar, R. A. Heckemann, J. V. Hajnal, and A. Hammers. Diffeomorphic registration using B-splines. In *Medical Image Computing and Computer-Assisted Intervention*, volume LNCS 4191, pages 702–709. Springer, 2006.

[10] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of cognitive neuroscience*, 3(1):71–86, 1991.

[11] J. S. Uleman, S. A. Saribay, and C. M. Gonzalez. Spontaneous inferences, implicit impressions, and implicit theories. *Annual Reviews of Psychology*, 59:329–360, 2008.

[12] A. Vinciarelli and G. Mohammadi. A survey of personality computing. *IEEE Transactions on Affective Computing (to appear)*, 2014.

[13] P. Viola and M. J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.

[14] A. Wright. Current directions in personality science and the potential for advances through computing. *IEEE Transactions on Affective Computing (to appear)*, 2014.