Recognition of Personality Traits from Human Spoken Conversations

A. V. Ivanov[1], *G. Riccardi*[1], *A. J. Sporka*[2] *and J. Franc*[3]

[1]Department of Information Engineering and Computer Science, University of Trento, Italy
[2]Dept. of Computer Graphics and Interaction, Faculty of Electrical Engineering, CTU[†],
[3]Dept. of Psychology, Faculty of Arts, Charles University, Prague, Czech Republic[†]
ivanov@disi.unitn.it, riccardi@disi.unitn.it, sporkaa@fel.cvut.cz, jakub.franc@gmail.com

Abstract

We are interested in understanding human personality and its manifestations in human interactions. The automatic analysis of such personality traits in natural conversation is quite complex due to the user-profiled corpora acquisition, annotation task and multidimensional modeling. While in the experimental psychology research this topic has been addressed extensively, speech and language scientists have recently engaged in limited experiments. In this paper we describe an automated system for speaker-independent personality prediction in the context of human-human spoken conversations. The evaluation of such system is carried out on the PersIA human-human spoken dialog corpus annotated with user self-assessments of the Big-Five personality traits. The personality predictor has been trained on paralinguistic features and its evaluation on five personality traits shows encouraging results for the conscientiousness and extroversion labels.

Index Terms: Automated personality prediction from speech, human–human dialog analysis.

1. Introduction

Understanding human personality and its manifestations in human behavior has been a long term goal of experimental psychology and more recently of researchers in the human-machine interaction and behavioral analytics. Such understanding would facilitate the human-machine spoken interaction design. The machine would have a possibility to customize its linguistic and behavioral patterns according to the expected needs of the human counterpart. Nass [15] as well as Bickmore [4] suggest that matching users' personality increases the efficiency of the interaction and is more natural for the user. In the experimental psychology research this topic has been addressed extensively [14], while speech and language scientists have recently engaged in promising experiments [13, 17]. There have been studies on how author's personality affects the particular style of the short textual communications (e-mails, blog entries) [9], the choice of particular parts of speech [16]. A comparison of the role of linguistic cues in spoken and textual communication is found in [13]. To this date the technology of personality assessment from speech (especially from paralinguistic cues) has had little attention from the research community. Recently in [17, 18] the authors designed a controlled experiment to acquire user-profiled speech utterances. However such corpus is collected by enacting different personality trait values (e.g extrovert vs introvert) using a single professional actor. Such speaker reads a given paragraph and produces relatively short ($\approx$ 20 sec.) utterances. The analysis of such personality traits corpora from natural conversations is quite difficult due to the user-profiled corpora acquisition, and also its annotation task and multidimensional observations. There are traditionally two types of personality trait assessment. The first is a self-assessment from the user under study and the second from a domain expert that evaluates users' traits. Last but not least personality may manifest in language-specific (e.g. spoken words or lexical statistics) or non-verbal cues [21]. In this paper we describe an automated system for speaker-independent personality prediction from human-human spoken conversations. The evaluation of such system is carried out on the PersIA human-human spoken dialog corpus annotated with user self-assessments of the Big-Five personality traits. The paper is organized as follows: second section discusses personality metric which was chosen for the experiment; third is devoted to the description of the PersIA human–human dialog corpus data collection process; section four describes the classification system and the personality trait prediction experiments.

[†] This is the current affiliation. The work has been done while collaborating with University of Trento.

2. Big-Five Personality Traits

There are numerous theories of personality as well as there are numerous definitions of personality, representing different views of the human beings and their behavior. Most frequently used in the computation-related psychological literature are the trait models (theories) of personality. According to [1], the personality traits are *"enduring patterns of perceiving, relating to, and thinking about the environment and oneself that are exhibited in a wide range of social and personal contexts"*. The traits are considered the features that are relatively stable over time and are assumed to affect the behavior of the individual. A personality is described through traits from a predefined set. The behavior of a person may be explained by a combination of the traits [20].

The Big Five model [5] is generally the most used of the trait personality models. It describes the human personality as a vector of five values corresponding to bipolar traits:

- Openness to experience: A preference to a varying experience, an appreciation for art, emotion, adventure, etc.
- Conscientiousness: A tendency to have a planned behavior (as opposed to spontaneous responses), a manifestation of self-discipline.
- Extroversion: "Energetic" behavior, an outgoing attitude, seeking the company of others.
- Agreeableness: Compassion and cooperativeness (as opposed to suspicion)

- Neuroticism: A tendency to "mood swings", a tendency to negative emotions such as anger or vulnerability.

This model is a popular choice among language and computer science researchers. It has been used as a framework for personality identification as well as simulation. Mairesse et al. [13] present a study of correlation of linguistic features on different levels (prosody, word choice, syntax) and the components of the Big-Five personality vector. Argamon et al. [3] demonstrate that it is possible to determine extroversion and neuroticism of an individual from a sample of informal text written by that person. Zen et al. [21] describe a system that is capable of recognition of extroversion and neuroticism from visual surveillance.

The model may be employed also in the other direction, that is for the synthesis of text or speech that perspires a desired personality. Mairesse et al. [12] presents PERSONAGE, a framework for generating the restaurant recommendations in textual form. André et al. [2] are using several Agreeableness and Extroversion of the Big Five model to simulate conversation of virtual agents in which they present different cars for sale to the observer of this dialog. The Big-Five model is used also in other interactive applications, e.g. [7] describes a model that helps consistently diversify the behavior of members of a simulated crowd.

3. The Data

3.1. Simulated Tourist Call Center

In this paper we experiment with the PersIA corpus that was collected during a controlled study of the effects of the personality on the user experience [6] of a simulated human-operated tourist call center. The purpose of this data collection was to gather linguistic as well as acoustic data so that insights to the manifestation of various personalities could be studied. The motivation for this effort was to gather necessary knowledge to build "**Pers**onable and **I**ntelligent virtual **A**gents", hence the name of the corpus.

The corpus contains a series of human–human conversations. The simulation of the tourist call center was realized by means of role-playing. Two separate groups of participants were assembled: The Users and the Agents. Each participant's personality has been measured by a Big Five personality test [10], translated into Italian with a back-translation verification.

Each User was expected to make a series of telephone inquiries and each Agent – provide relevant answers in order to fulfill the tourist tasks. Each User was given a schedule of calls to make and a task to accomplish during each call. The tasks included open-ended tasks (e.g. "Find out what you can do in Bolzano."), simple look-up tasks (e.g. "Find out the screening time of The Unborn movie in Trento tonight."), look-up tasks where an alternative must be negotiated (similar to previous, but no solution could be found without loosening the constraints, such as suggesting another town or time), and tasks for which there was no solution (e.g. "Find out the location of the mosque in Trento.")

The schedule and the task assignments were communicated to the Users through a dedicated web site. Immediately after the call, the Users were asked to compile a questionnaire to assess the user experience on this web site, as shown in Figure 1. The user experience was quantified by three indexes, Effectiveness, Efficiency, and Satisfaction [1].

[1] in accordance with the ISO 9241 standard

Each Agent was assigned a three-hour slot during which he or she was supposed to take calls from the Users. Each Agent was provided with a sheet of paper that contained a summary of knowledge that the simulated agency was supposed to provide. The information contained in the sheet was fictive; however, a care was taken to assemble the sheet so that it contained believable facts.

The Agents were using a dedicated software tool to handle all telephone calls. A simple protocol was established to identify the callers: Each call started with a few prompts, triggered by the Agent, which asked the User to state his or her identification number and ID number of the task. The tool automatically recorded the speech in separate channels together with the identification numbers of the User and the Agent and the task ID number. The training of the Agents was minimal: only to ensure that each Agent was able to handle the calls properly.

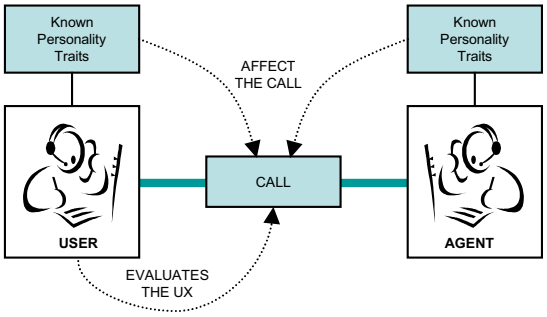

Figure 1: Overview of the study organization.

3.2. Corpus Description

A total of 24 speakers took part in the experiment.12 Users (mean age 22 y.o.; age standard deviation 6.0; 7 males, 5 females) and 12 Agents (mean age 27 y.o.; age standard deviation 6.5; 6 males, 6 females) took part in the test. The speakers were recruited from the university staff and students.

The Big Five profiles of all the speakers were known. Each grade of personality trait was converted into a binary label "L" for low-grade and "H" for high-grade of a given trait. It was done according to a median split of the score population, resulted from the personality self-assessment questionnaire during the data collection. Each speaker was recorded in a separate channel. The signal from each channel has been segmented into utterances and transcribed.

Out of 144 calls that were possible (each User $\times$ each Agent) 119 calls took place. The total duration was 2 hours and 14 minutes. The transcribed corpus contains approx. 15,000 words. Although there was a significant variation in the duration of the individual dialogs, the mean dialog duration was about 70 seconds.

Due to the asymmetry of the roles, the amount of speech recorded and the structure of utterances in terms of dialog act analysis were different for the Users and for the Agents. While there were 6,145 tokens produced by the Users and 9,371 tokens by the Agents. For the experiment described in this paper only the Agents' speech was used (Agent subcorpus).

The entire corpus was processed with TreeTagger in order to obtain a part-of-speech tag for each word. For each trait, we compared the POS distribution of the "L" and "H" level of that trait. The χ^2-tests, one for each such a comparison, revealed a significant pairwise difference between the "L" and "H" dis-

tributions ($\chi^2(16, N = 9371) > 68, p << 0.0001$ for all tests). Such statistical distributions motivate future work on the linguistic manifestation of personalities.

4. Experiment

In order to explore a possibility to assess personality from a dialog recording, a state-of-the-art emotion recognition system has been trained and tested on the data, coming from the PersIA database. The system consisted of an openSMILE-based feature extraction [8] and the boostexter classifier [19].

4.1. Feature Extraction

For each of the whole set of 119 dialogs we extracted the speakers' channel (user and agent). Each channel was processed to remove long silence segments. This was done with the help of Automated Turn Segmenter [11]. PersIA data collection has resulted in different amount of the spoken evidence between the information providing agent and the user requesting assistance (see Table 1 for details). The user channel typically contains a few short questions and a lot of back-channel activity, while the agent's spoken activity is much longer and more diverse in content. Only the agent channels were considered for further processing.

Table 1: *PersIA corpus speech statistics for users and agents.*

Speech Mean Duration	User	Agent
Seconds	22.23	36.01
Tokens	54.44	86.22
Dialogs (per Agent)	N of Dialogs	Total, sec
Agent 1	10	563
Agent 2	8	376
Agent 3	10	256
Agent 4	11	172
Agent 5	12	446
Agent 6	10	284
Agent 7	9	261
Agent 8	8	227
Agent 9	12	589
Agent 10	12	506
Agent 11	8	235
Agent 12	9	391

The data was split 12-ways for cross-validation by leaving one speaker out (LOSO). Thus each of the 12 test sets contained the data coming from a single speaker that was not present in the corresponding training set. This splitting strategy is essential as it was observed that random splitting of the data leads to the classifier memorising individual voices and the associated personality label, which results in the much higher average performance expected in cross-validation.

Feature extraction was performed with the predefined openSMILE `emo_large.conf` feature set [8]. The whole set consisted of a detailed statistical description of the basic speech features. Those basic features have included 13 MFCC coefficients; envelopes in the individual mel-frequency channel (26 values); signal log-energy in a sliding window; band envelope for bands 0-250 Hz, 0-650 Hz, 250-650 Hz, 1-4 kHz; spectral centroid and flux; spectral rolloff to the levels 25%, 50%, 75%, 90%; position of the spectral maximum and minimum; pitch estimate; zero crossing rate. Additionally the first and the second derivatives of each of the mentioned values were estimated with a standard non-causal FIR approximation, thus giving 168 individual dynamically-changing feature readings. The statistics were taken uniformly over the entire spoken content of the corresponding channel. The statistical parameters, which were estimated for each individual feature reading, have included dynamic range; centroid; standard deviation; skewness and kurtosis; position of the maximum and minimum values; distance between minimum; maximum and the mean; coefficients of linear and quadratic regression; linear and quadratic errors of regression; quartile and percentile analysis for 95% and 98%; interquartile ranges; zero crossing rate; number of individual peaks; average inter-peak distance; arithmetic mean of peak values; number of non-zero values; arithmetic, quadratic and geometric means for all and only non-zero values. Thus, the final feature vector consisted of 6552 individual real-valued parameters.

4.2. Classification Experiment

A separate classifier was trained for each cross-validation fold. A blind stop after 500 iterations was chosen as a stopping criterion, thus, avoiding tuning the performance on the test set. Test results of each of the individual classifiers were summed up to produce a final statistics. In the classification experiment a binary label ("L" or "H", see section 3.2 for details) was used. A binarized self-assessment score was used as a supervision signal for the training pairs of the classifier. Although expert and self-assessment derived scores are found to be correlated with each other, the "self-assessment report"-derived labels may lead to a lower performance [14].

Table 2: *Performance of predictors of individual Big Five personality traits (Openness, Conscientiousness, Extroversion, Agreeableness, Neuroticism). "CORR" – number of correctly labeled dialog channels (the total number of dialog channels was always 119), "Acc. %" – recognition accuracy in percents,"Chance %" – performance, corresponding to the random drawing from the prior distribution of labels,"p-value" – probability to observe at least the experimental ratio of correct guesses after random drawing from the prior distribution of labels.*

Personality Trait	CORR	Acc. %	Chance %	p-value
Openness	48	40.34	52.97	0.9962
Conscientiousness	**113**	**94.96**	**73.17**	$\mathbf{9.8 \cdot 10^{-11}}$
Extroversion	**75**	**63.03**	**50.00**	$\mathbf{1.6 \cdot 10^{-3}}$
Agreeableness	67	56.30	54.83	0.3401
Neuroticism	39	32.77	50.00	0.9999

Table 2 summarizes the classifier accuracy over the five personality traits. The performance for conscientiousness and extroversion significantly rises above the chance level. The probability to observe a result, better than the experimental (i.e. *at least* "Acc %"-ratio of sucessfully guessed personality labels) while randomly drawing from the prior distribution of labels is computed with a binomial test. In the case of conscientiousness and extroversion this probability is significantly less than the standard significance level of $p = 0.05$. This fact supports the suggestion, that spoken evidence may be used for detection of the mentioned personality traits (conscientiousness and extroversion).

Table 3: *Performance of the predictors of extroversion with some intermediate cases being removed from the consideration. "RM scores" – particular scores being removed, "Total" – the total number of dialog channels."CORR", "Acc. %", "Chance %", "p-value" have the same meaning as in Table 2.*

RM scores	CORR	Total	Acc. %	Chance %	p-value
6	75	108	69.44	50.43	$2.0 \cdot 10^{-5}$
6, 7	63	87	72.41	56.35	$6.8 \cdot 10^{-4}$

Figures 2 and 3 are reflecting the distribution of the system predictions through the trait scores of the agents. It is evident

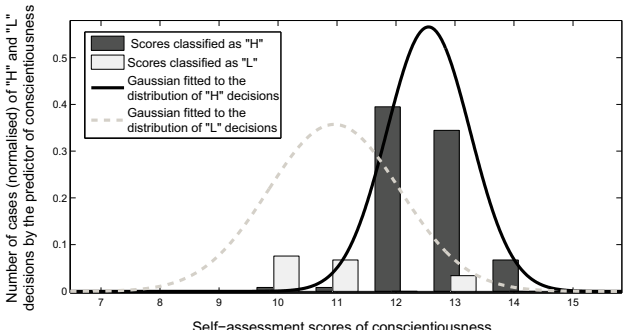

Figure 2: *Prediction of the degree of conscientiousness from paralinguistic features.*

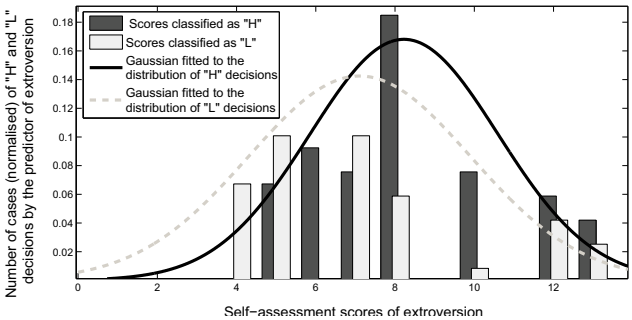

Figure 3: *Prediction of the degree of extroversion from paralinguistic features.*

that for conscientiousness and (to a somewhat lesser extent) for extroversion there is a good separation between the score ranges, which correspond to the different system's predictions. As one can see in Fig. 3, a significant portion of the erroneous prediction occurs when extroversion trait score has an intermediate level. To assess the performance over the extreme (high and low scores) extroversion axis we withheld the intermediate scores from the test. Table 3 shows results of withholding speaker who scored 6 and 7 on extroversion scale.

5. Conclusion

The analysis of human behavior has been addressed in the context of personality's manifestations in language. We have investigated how machines can be trained to automatically predict from observations over the duration of a conversation of such personality traits. Such prediction has been based on speaker independent models and incorporates paralinguistic features only. This corpus has been specifically designed to study the personality-related aspects of natural human-human verbal dialog communication. We have achieved statistically significant performances in predicting some of the speaker personality traits, namely the level of conscientiousness and extroversion.

6. Acknowledgments

This work was partially supported by the European Commission Marie Curie Excellence Grant for the ADAMACH project (contract No. 022593), the Livememories project funded by Autonomous Province of Trento and MSMT Czech Republic under the research program LC-06008 (Center for Computer Graphics).

7. References

[1] American Psychiatric Association. Diagnostic and statistical manual of mental disorders, dsm-iv, 2000.

[2] E. André, T. Rist, S. van Mulken, M. Klesen, and S. Baldes. *Embodied Conversational Agents*, chapter Automated Design of Believable Dialogues for Animated Presentation Teams., pages 220–255. MIT Press, 2000.

[3] S. Argamon, S. Dhawle, M. Koppel, and J. W. Pennebaker. Lexical predictors of personality type. In *Proc. of the Joint Ann. Meeting ofInterface and the Classif. Soc. of N. America*, 2005.

[4] T.W. Bickmore and R.W. Picard. Establishing and maintaining long-term human-computer relationships. *ACM Trans. Comput.-Hum. Interact.*, 12:293–327, June 2005.

[5] J. Digman. Personality structure: emergence of the five factor model. *Annual Review of Psychology*, 41:417–40, 1990.

[6] A. Dix, J. Finlay, G.D. Abowd, and R. Beale. *Human–Computer Interaction*. Prentice-Hall, 3rd edition, 2004.

[7] F. Durupinar, J.M. Allbeck, N. Pelechano, and N.I. Badler. Creating crowd variation with the ocean personality model. In Lin Padgham, David C. Parkes, Jrg Mller, and Simon Parsons, editors, *AAMAS (3)*, pages 1217–1220. IFAAMAS, 2008.

[8] F. Eyben, M. Wöllmer, and B. Schuller. openSMILE - the Munich versatile and fast open-source audio feature extractor. In *Proc. ACM Multimedia (MM), Florence, Italy*, pages 1459–1462, 2010.

[9] A. J. Gill and R. M. French. Level of representation and semantic distance: Rating author presonality from texts. In *Proc. of the 2nd European Cognitive Science Conference (EuroCogsci07)*, 2007.

[10] S.D. Gosling, P.J. Rentfrow, and W.B. Swann. A very brief measure of the big-five. *J. of Res. in Personality*, 37:504–528, 2003.

[11] A. V. Ivanov and G. Riccardi. Automatic turn segmentation in spoken conversations. In *Proc. of Interspeech'2010, Makuhari, Japan*, 2010.

[12] F. Mairesse and M. Walker. PERSONAGE: Personality generation for dialogue. In *Proc. 45th Meeting of ACL*, pages 496–503, 2007.

[13] F. Mairesse, M. A. Walker, M. R. Mehl, and R. K. Moore. Using linguistic cues for the automatic recognition of personality in conversation and text. *J. Art. Intelligence Res.*, 30:457–500, 2007.

[14] M. R. Mehl, S. D. Gosling, and J. W. Pennebaker. Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *J. of Pers. and Soc. Psych., 90*, pages 862–877, 2006.

[15] C. Nass and S. Brave. *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship*. MIT Press, 2005.

[16] J. Oberlander and A. J. Gill. Individual differences and implicit language: Personality, parts-of-speech and pervasiveness. In *Proc. of the 26th Annual Conference of the Cognitive Science Society, Chicago, IL, USA*, 2004.

[17] T. Polzehl, S. Moller, and F. Metze. Automatically assessing acoustic manifestations of personality in speech. In *Spoken Language Technology Workshop, 2010 IEEE*, pages 7 –12, 2010.

[18] T. Polzehl, S. Moller, and F. Metze. Automatically assessing personality from speech. In *Proceedings of the 2010 IEEE Fourth International Conference on Semantic Computing*, ICSC '10, pages 134–140, Washington, DC, USA, 2010. IEEE Computer Society.

[19] R. Schapire and Y. Singer. Boostexter: A boosting–based system for text categorization. In *Mach. Learn.*, pages 135–168, 2000.

[20] J. Trouvain, S. Schmidt, M. Schmitz, and W. J. Barry. Modeling personality features by changing prosody in synthetic speech. In *Proc. Speech Prosody 2006, Dresden, Germany*, page Paper 088. Intl. Speech Communication Association, 2006.

[21] G. Zen, B. Lepri, E. Ricci, and O. Lanz. Space Speaks – Towards Socially and Personality Aware Visual Surveillance. In *Proceedings of the 1st ACM international workshop on Multimodal pervasive video analysis*, pages 37–42. ACM, 2010.