2011 IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing

Contextual Modeling of *Personality States*' Dynamics in Face-to-Face Interactions

Jacopo Staiano*, Bruno Lepri†‡, Kyriaki Kalimeri†, Nicu Sebe* and Fabio Pianesi†
*Department of Information Engineering and Computer Science
University of Trento, Italy
{staiano,sebe}@disi.unitn.it
†Fondazione Bruno Kessler, Trento, Italy
{lepri,kalimeri,pianesi}@fbk.eu
‡Massachussets Institute of Technology, Boston, USA

***Abstract*—In this paper, we investigate the effectiveness of several dynamic graphical models for the task of *personality states* classification in a meeting scenario. *Personality states* are defined in social psychology literature as specific behavioral episodes in which a given subject exposes behaviors connected to a certain personality trait. The *personality states* we are addressing are those corresponding to the Big Five traits.**

I. Introduction

In everyday life, people constantly and unconsciously exploit the ability to describe others on the basis of their behaviors: for instance, descriptors such as *talkative, bold, sociable* refer to the *Extroversion* personality trait dimension; *responsible, attentive*, refer to the *Conscientiousness* personality trait, and so on. This ability of describing others through personality descriptors is crucial for explanation and prediction of others' behaviors. The importance of personality has been acknowledged in a (increasing) number of studies: it has been showed that the basic dimensions of adaptivity [6] as well as people's attitude towards computers [16] and conversational agents [15], are influenced by personality traits. In the social networking context, following analysis of text messages, personality matching between users has been proved to increase the chances of successful relationships [3]. In general, finding means to automatically obtain information about people's personalities is considered very significant in order to let machines act in a *proactive* fashion or endowing them with the folk-psychological capability of explaining/predicting people's behaviors [1]. Several works have explored automated personality analysis [10], [12], [9], often targeting the Big Five model of personality [2]. All these approaches to the automatic recognition of personality have more or less implicitly adopted the so-called person perspective on personality [5]: for a given behavioral sample, classify whether the sample belongs to an extrovert or introvert (or equivalently, to a neurotic or an emotionally stable, and so on). The problem with this approach is that it assumes a direct and stable relationship between, e.g., being extravert and acting extravertedly (e.g., speaking loudly, being talkative, etc.). Extraverts, on the contrary, can often be silent and reflexive and not talkative at all, while introverts can at time exhibit extraverted behaviors. While the person perspective has often dismissed these fluctuations of actual behavior as statistical noise, it has been suggested by Fleeson [5] that they are meaningful. The social psychology literature has coined the term *personality states* to refer to concrete behaviors that can be described as having a similar content to the corresponding personality traits. In other words, a personality state describes a specific behavioral episode wherein a person behaves more or less, introvertly/extravertly, more or less neurotically, etc. Personality can then be reconstructed through density distributions over personality states, with parameters such as means, standard deviations, etc., summarizing what is specific to the given individual.

In this paper we address the automatic classification of *personality states*. Hence, we will be concerned with classifying whether people behaved extravertly or introvertly; neurotically or in an emotionally stable manner; and so on. To this end, we exploit graphical dynamic models that may explicitly incorporate hypotheses about the relationships among personality, actual behavior of the target and situational aspects. In particular, we will hypothesize that: a) personality states are the only causal determinants of people behavior; b) people behavior is modulated by the behavior and the personality states of their meeting parties (situational factor). The importance of those relationships will be investigated by comparing the performance of three sequential models: generative linear-chain Hidden Markov Models (L-HMM) and discriminative linear-chain Conditional Random Fields (L-CRF) incorporating only hypothesis (a); generative Influence Modeling (IM) incorporating (a) and (b).

II. Corpus

Our experiments were performed on a 4-meetings subset of the Mission Survival II corpus [12]; the video stream corresponding to each meeting participant was split into 5-minute long slices. Personality state annotation was performed by volunteers using the Ten Item Personality Inventory [7], a 10-item questionnaire developed to obtain a quick measure of the Big Five dimensions. Annotators were required to assess on a 1 to 7 scale the personality of the participants based on the 5-minute slices containing a close-up view of the subject with

Bruno Lepri's research was supported by PERSI project funded by Marie Curie - COFUND - 7th Framework.

978-0-7695-4578-3/11 $26.00 © 2011 IEEE
DOI

the synchronized audio. Contextual information (the behavior of the other meeting participants) was available only through the audio channel. Each video was annotated by three different annotators, and each annotator saw a given subject no more than once. The global means and standard deviations computed from the annotated personality states amount respectively to [4.18-1.66] for Extraversion, [4.55-1.05] for Agreeableness, [4.99-1.12] for Conscientiousness, [5.07-1.04] for Emotional Stability, and [3.96-1.25] for Openness.

In order to perform the classification experiments, scores were quantized (Low/High) for each personality state according to the following procedure: for each subject and each slice, judges' score were assigned label L(ow) or H(igh) according to whether they were lower or greater than 4. The majority vote on the three resulting binary scores was then taken as the (subject-slice) score on that dimension. Using this procedure, the obtained distribution of high-low values in the dataset are, respectively: [251-245] for Extraversion, [175-321] for Agreeableness, [139-357] for Conscientiousness, [112-384] for Emotional Stability, and [310-186] for Openness. It is evident that Agreeableness, Emotional Stability and Conscientiousness are strongly unbalanced towards the High class, whereas with Extraversion and Openness the unbalance reduces or reverses. This strong unbalance may be due to our particular scenario: during meetings neurotic, disagreeable and unashamedly negligent behavioral manifestations are not usual.

Our subjects showed a good degree of consistency in their behaviors: if a person was behaving extravertly/introvertly at time t he/she was more likely to continue behave the same way at $t+1$ (respectively, transition probabilities of 0.64 and 0.65). Some exceptions in this regard are (a) the transition from a low conscientiousness state to an high conscientiousness state (0.62): and (b) the transition from an emotionally stable behavior to a neurotic behavior (0.67). Finally, a transition from a disagreeable behavior to an agreeable behavior (0.54) is as likely as continuing with a disagreeable behavior (0.46).

III. Feature Extraction

A total of 37 features, including both low- and high-level features, were automatically extracted from the meeting corpus.

A. Low-level features

We focused on two classes of features: *Activity* and *Emphasis* [11], measuring vocal signals in social interactions. *Activity*, implying conversational activity level, usually indicates interest and excitement. It is measured by the z-scored percentage of speaking time (mean and standard deviation of energy per frame, average length in seconds of voiced segments and of speaking segments, fraction of spoken time and voicing rate). For this purpose, the speech stream of each participant is first segmented into voiced and non-voiced segments, and then split into speaking and non-speaking segments. *Emphasis* is often considered a signal indicating the strength of the speakers motivation. Moreover, the variability of emphasis may signal an openness to influence from other people. *Emphasis* is measured by the variation in prosody, i.e. pitch and amplitude. For each voiced segment, the mean energy, frequency of the fundamental formant and the spectral entropy are extracted (mean of formant frequency, confidence in formant frequency, spectral entropy, the value of the largest autocorrelation peak, location of the largest autocorrelation peak, the number of the largest autocorrelation peak, the time derivative of energy in frame). The mean-scaled standard deviation of these extracted values is then estimated by averaging over longer time-periods.

B. High-level features

Social attention features were extracted by jointly processing the audio-video channels using the hierarchical approach developed in [9]: the output of a Cylindrical Head Model [20] head-pose tracker was used as a proxy for the subject's gaze and finely tuned with the output of a sub-pixel accurate visual-gaze estimation system [19]. Social attention features were defined as the outcome of such joint processing of head-pose and visual-gaze: for each subject, in every frame, *Attention Given* (to at least one of the other participants) and *Attention Received* (by at least one of the other participants) were thus available. The audio channel was processed through use of a Voice Activity Detection system based on the long-term spectral divergence algorithm detailed in [14]. Speaking time was calculated as the percentage of frames in which voice activity was detected over the duration of the processed slice. For every participant p, the audio-visual features were then fused in order to obtain *Attention Given While Speaking*, *Attention Given While Not Speaking*, *Attention Received While Speaking*, and *Attention Received While Not Speaking*.

IV. Feature selection

Correlation-based feature selection is a subset selection technique whose objective is to determine an optimal subset of features [8]. This method evaluates the merit of a subset of features computing the individual predictive ability of each feature along with the degree of redundancy among them. The preferred and selected features are those highly correlated with the target value and with low inter-correlation values. This method is used in conjunction with the Best First search strategy (in our case a forward search) that makes a search in the space of features subsets, using a greedy hill-climbing with a backtracking facility. The selection was based on 10-fold cross validation; features selected in at least 7 folds were chosen to train and test the models. The features selected for each personality trait as follows: mean of attention given, attention received, attention received while not speaking, formant frequency, spectral entropy, energy in frame for *Extraversion*; mean of formant frequency, standard deviation of spectral entropy for *Neuroticism*; mean of formant frequency, confidence in formant frequency, standard deviation of spectral entropy, number of autocorrelation peaks for *Agreableness*; mean of attention received, attention received while speaking, attention received while not speaking, formant frequency; standard deviation of spectral entropy for *Openness*; mean of attention received, attention given while speaking, attention received while not speaking, standard deviation of spectral entropy, value of the largest autocorrelation peak for *Consciousness*.

V. Modeling personality states dynamics using graphical models

As discussed previously, there are certain regularities in personality state change that sequential models might be able to capture: all classifiers (generative L-HMM and IM, discriminative L-CRF) were exploited to this end. In addition, IM assumes that people influence each other and that the current personality state of a person is influenced by the personality states of the other participants, besides being influenced by the persons previous personality states.

Linear-chain Hidden Markov Models consider the temporal relationships between the samples and define the prior probability of the classes in the current sample as depending upon the posterior probabilities of the classes in the last sample [13]. More precisely, a left-to-right HMM model for each personality dimension was represented as follows: t, time; $y(t)$, the feature vector; $x(t)$, the personality state; $p(x)$, the priors for the personality states; $p(x(t)|x(t-1))$, the personality states transitions probabilities; $p(y(t)|x(t))$, the conditional distribution of the observed feature vector given the current personality state at time t. We assumed subjects independence and the feature sequences (one per subject) from all the four meetings were used to train a single HMM. The training was done using the standard Expectation Maximization (EM) algorithm. For prediction, each person is represented by an independent instantiation of the same Markov process. Thus, four independent HMMs represent the four people in a meeting. For classification, we used the standard Viterbi algorithm to compute the most likely sequence of personality states.

Linear-chain CRFs are undirected graphical models that compactly represent the conditional probability of a particular label sequence, Y, given a sequence of observations X. Modeling the conditional probability of the label sequence $P(Y|X)$ rather than the joint probability of both the labels and observations $P(X,Y)$, as done by hidden Markov models, allows CRFs not to require any independence assumptions between the observation variables in X for tractable inference. Model parameters can be learned using an iterative gradient method such as BFGS method. Inference can be performed using a slightly altered version of the Viterbi algorithm [18]. As for linear-chain HMMs, we assumed subject independence and each left-to-right CRF represents a different meeting participant in our setting. In the linear chain example, the $x(t)$ items correspond to the features and the $y(t)$ to labels to be predicted at time t, for example speech activity and social attention features, and personality states, respectively. In order to run our experiments we employed the GRMM (Graphical Models in Mallet) implementation [17].

In order to model the co-temporal relationship between the personality states played by the four meeting participants, we used the Influence Model approach [4]. The representation of the model is similar to the HMMs with a small difference. Each Markov process independently is non-stationary and the transition probabilities $p\left(x_i\left(t\right)|x_i\left(t-1\right)\right)$ for a chain i is given as

$$p\left(x_i\left(t\right)|x_i\left(t-1\right)\right)=\sum_{j=1}^{C_S}\left(d_{j,i}\sum_{x_j=1}^{x_N}a\left(x_j,x_i\right)p\left(x_j\left(t\right)\right)\right)$$

where $d_{j,i}$ represents the influence between processes j and i, and $a\left(x_j,x_i\right)$ represents the influence between the states $x_j$and $x_i$of the interacting processes j and i. For each personality dimension, we used four interacting Markov processes to model the evolution of the personality states of the four meeting participants. The observations for the individual processes are the participants raw features. The latent states for the individual processes are the personality states labels (high and low for each trait). In the training phase of influence modeling, we find out the observation statistics of two classes personality states (high and low), as well as the interaction of different participants with the Expectation Maximization (EM) algorithm. In the prediction phase, we infer the individual participants personality states (extroverts vs introverts, neurotics vs emotionally stable) based on observations about her/his behavior, as well as on observations about the interactions with other participants.

TABLE I
Experimental Results, Average F1 Scores

	Extr.	Agr.	Consc.	EmSt.	Op.	(marg)
L-C HMM	0.56	0.50	0.57	0.49	0.59	0.54
L-C CRF	0.63	0.43	0.39	0.48	0.58	0.50
IM	0.54	0.50	0.58	0.54	0.51	0.53
(marg)	0.57	0.46	0.50	0.49	0.56	

VI. Experimental Results

Table I reports the performance results for the three graphical models we used in our experiments; each model was evaluated following a leave-one-meeting-out cross-validation strategy.

As can be seen from the column-wise marginal values in Table I, the overall performances of the three classifiers fall in range from 0.46 to 0.57 and differ little.

HMM-based methods obtain higher average F1 values on Agreeableness, Emotional Stability and Conscientiousness, while the two types of methods have similar performances on Openness and Extraversion. Concerning the role of context, the comparison between methods that exploit it (IM) and those that do not (L-CRF and L-HMM) does not reveal substantial differences on Agreeableness, Emotional Stability and Conscientiousness. Non-contextual methods, in turn, slightly improve their performance over the other two methods on Openness and Extraversion (see Fig. 1).

A distinction seems to emerge between Agreeableness, Emotional Stability and Conscientiousness personality states, on one hand, and Extraversion and Openness, on the other. The former seem to be better analyzed by HMM-based methods,

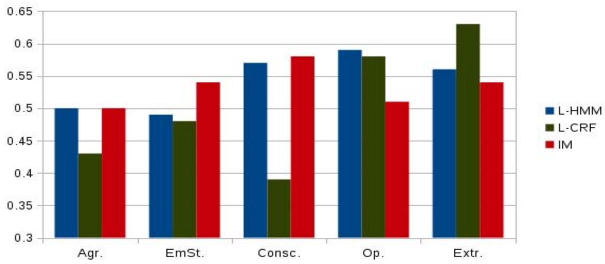

Fig. 1. Average F1 scores for L-HMM, L-CRF, and IM classifiers.

TABLE II
PEARSON RESIDUAL FOR THE BEST PERFORMING CLASSIFIERS

	Extr.	Agr.	Consc.	EmSt.	Op.
	L-CRF	L-HMM	IM	IM	L-HMM
Low	**3.91**	1.31	**3.97**	0.39	1.55
High	**4.34**	1.55	**3.19**	**4.6**	**5.94**

whereas the latter seem more amenable to methods that exploit contextual information.

One possible consequence of the strong unbalance between classes is that classifiers do actually learn little, if any, of the data structure and only end up working on the basis of prior probabilities. To check whether this was the case, we compare the best results in Table I for each personality state to the expected results of a classifier that uses prior probabilities for each of them. The comparison will be conducted in terms of Pearson residual computed on the relevant confusion matrices. We will consider only Pearson residuals for hit classes, whose interpretation is straightforward: negative values indicate that the number of hits in the relevant class is lower than that of the baseline classifier; positive values indicate the converse. We exploit the N(0, 1) distribution of Pearson residual and fix a 2.5 threshold for significant (in/de)crease with respect to the baseline. The results are reported in Table II; bold figures mark significant results.

Only in two cases out of five, Extraversion with L-CRF and Conscientiousness with IM, there was a significant improvement over the baseline in both classes. For two other personality states, Emotional Stability and Openness, the corresponding classifier (IM and L-HMM, respectively) improves the results only in the High class; L-HMM performance on Openness is to be noticed since it is obtained in the less frequent class. Finally, L-HMM performance with Agreeableness is not statistically different from those of the baseline.

VII. CONCLUSIONS

To take stock of the data discussed in the experiments, there is no evidence of a global advantage in using contextual information for getting at personality states. More detailed analysis, however, suggests that the Influence Model is a possible candidate for improved performance, at least on states like Conscientiousness and, possibly, Emotional Stability. Other states, such as Extraversion and Openness, in turn, seem less likely to benefit from contextual information, suggesting that the modulation effect of the other groups member behavior is limited. A final cautionary note is still in order: all the data and results discussed have been obtained from an often strongly unbalance corpus. Though the unbalance is not due to chance, but to the very nature of the social situation our subjects were involved in, it might well have prevented a full exploitation of the power of the classifiers employed. Finally, our results are obviously affected by the assortment of feature used. Future work should focus on larger data collections, with a finer grained definition of states and larger and more state-specific arrays of features. With all these caveats, this is one of the first works to address the automatic classification of personality states, a step that we have argued to be crucial to approach the personality-behavior problem in an effective way.

REFERENCES

[1] K. Andrews. It's in your nature: A pluralistic folk psychology. *Synthese*, 165(1), 2008.
[2] P. T. Costa and R. R. McCrae. Four ways five factors are basic. *Personality and Individual Differences*, 13(6):653–665, 1992.
[3] J. M. Dabbs, F. J. Bernieri, R. K. Strong, R. Campo, and R. Milun. Going on stage: Testosterone in greetings and meetings. *Journal of Research in Personality*, 35(1):27–40, 2001.
[4] W. Dong and A. Pentland. Modeling influence between experts. ICMI'06/IJCAI'07, pages 170–189, 2007.
[5] W. Fleeson. Toward a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology*, 80(6):1011–1027, 2001.
[6] D. Goren-Bar, I. Graziola, F. Pianesi, and M. Zancanaro. The influence of personality factors on visitor attitudes towards adaptivity dimensions for mobile museum guides. *UMUAI*, 16:31–62, 2006.
[7] S. Gosling. A very brief measure of the big-five personality domains. *Journal of Research in Personality*, 37(6):504–528, 2003.
[8] M. A. Hall. Correlation-based feature selection for discrete and numeric class machine learning. pages 359–366. Morgan Kaufmann, 2000.
[9] B. Lepri, R. Subramanian, K. Kalimeri, J. Staiano, F. Pianesi, and N. Sebe. Employing social gaze and speaking activity for automatic determination of the extraversion trait. ICMI-MLMI '10, 2010.
[10] 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 ARTIF INTELL RES.*, 30:457–501, 2007.
[11] A. Pentland. *Honest Signals: How They Shape Our World*. MIT Press, 2008.
[12] F. Pianesi, N. Mana, A. Cappelletti, B. Lepri, and M. Zancanaro. Multimodal recognition of personality traits in social interactions. ICMI '08, pages 53–60, 2008.
[13] L. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *P IEEE*, 77:257–286, 1989.
[14] J. Ramirez, J. C. Segura, C. Benitez, A. De La Torre, and A. Rubio. Efficient voice activity detection algorithms using long-term speech information. *Speech Communication*, 42:271–287, 2004.
[15] B. Reeves and C. Nass. *The media equation: how people treat computers, television, and new media like real people and places*. Cambridge University Press, 1996.
[16] J. F. Sigurdsson. Computer experience, attitudes toward computers and personality characteristics in psychology undergraduates. *Personality and Individual Differences*, 12(6):617–624, 1991.
[17] C. Sutton. Grmm: Graphical models in mallet. 2006.
[18] C. Sutton and A. Mccallum. *Introduction to Conditional Random Fields for Relational Learning*. MIT Press, 2006.
[19] R. Valenti, J. Staiano, N. Sebe, and T. Gevers. Webcam-based visual gaze estimation. ICIAP '09, pages 662–671, 2009.
[20] R. Valenti, Z. Yucel, and T. Gevers. Robustifying eye center localization by head pose cues. In *Proc. IEEE CVPR Conference*, 2009.