One of a Kind: Inferring Personality Impressions in Meetings

Oya Aran[1] and Daniel Gatica-Perez[1,2]
[1] Idiap Research Institute, Martigny, Switzerland
[2] Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland
(oaran,gatica)@idiap.ch

ABSTRACT

We present an analysis on personality prediction in small groups based on trait attributes from external observers. We use a rich set of automatically extracted audio-visual nonverbal features, including speaking turn, prosodic, visual activity, and visual focus of attention features. We also investigate whether the thin sliced impressions of external observers generalize to the whole meeting in the personality prediction task. Using ridge regression, we have analyzed both the regression and classification performance of personality prediction. Our experiments show that the extraversion trait can be predicted with high accuracy in a binary classification task and visual activity features give higher accuracies than audio ones. The highest accuracy for the extraversion trait, is 75%, obtained with a combination of audio-visual features. Openness to experience trait also has a significant accuracy, only when the whole meeting is used as the unit of processing.

Categories and Subject Descriptors

I.5.4 [**Pattern Recognition Applications**]: [Computer Vision, Signal Processing]; J.4 [**Computer Applications**]: Social and Behavioral Sciences—*Sociology, Psychology*

General Terms

Modeling, Algorithms, Theory

Keywords

Personality prediction, multimodal analysis, social interaction, nonverbal behavior

1. INTRODUCTION

There is an increasing interest in building machines capable of inferring and reacting to the individual traits of their users. One fundamental human trait is personality [1], and the Big Five model of personality has been adopted in psychology as one capable of explaining this construct from the highest level of abstraction. The Big Five model defines five clusters of personality dimensions: Extraversion (Ext), Agreeableness (Agr), Conscientiousness (Con), Emotional stability (Emo), and Openness to Experience (Ope).

The personality of individuals gets expressed in many everyday life situations, with family and friends and at work. The study of personality in the workplace and in group meetings in particular is relevant because the personality of team members (expressed through their nonverbal behavior and the content of their messages) can influence in important ways the group's decisions and actions [1]. The literature in nonverbal communication establishes that the study of personality in this context is complex for many reasons. First, groups involve multiple individuals interacting (each one playing a different group role or having a different degree of familiarity with one another). Second, the personality of an individual can be self-attributed or externally observed (so-called personality impressions), and it is known that these two types of measurements need not be the same. Third, personality is expressed through voice, face, body in the nonverbal channel, and also through the spoken words in non-unique forms. All these factors make the computational study of personality of small team members a challenging research subject for computing.

Among the attempts to model and infer personality traits with computational means, one of the initial works looked at small group meetings to predict two personality traits, extraversion and locus of control, using a large set of audio features and few visual features [2]. They used self-reported personality judgments as the ground truth to build their models. In [3], the authors investigated the automatic classification of the extraversion personality trait using speaking time and social attention behaviors as features. Judged personality impressions were used in [4] to model personality states in small group meetings with audio-visual nonverbal features. They use low level and high level speech features, and visual focus of attention features. Other works investigated personality prediction in other domains. In [5] and [6], the authors investigated a social media domain, vlogs, for the automatic analysis of personality using external observers' attributes. In [7], the authors estimated the Big Five traits judged by external observers on a dataset of broadcasted radio clips, which contain monologue-like presentations. Self presentations are used in [8] to predict self-reported personality traits. Some other works used sensors other than audio-visual ones to capture nonverbal behavior to predict personality. In [9], the authors processed the

ICMI'13, December 9–12, 2013, Sydney, Australia
Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-2129-7/13/12 ...$15.00.
http://dx.doi.org/10.1145/2522848.2522859 .

smartphone usage of people to predict self-reported personality traits. In [10], the authors used sociometers to capture the physical activity, speech activity, and interactions of people, and analyzed the correlations between the extracted features and the Big Five personality dimensions.

In this work, we have investigated the prediction of personality of individuals participating in small group discussions based on traits judged by external observers. Although most of the works in personality prediction have used self reported personality traits, judged personality impressions provide a different view of personality, as decoded by observers [11]. In our study, we use speaking turn and prosodic features as audio features and visual activity and visual focus of attention features as visual features. The visual activity features that are extracted from the participant provide information regarding the nonverbal body dynamism. In comparison to earlier work on personality prediction, which used either manually extracted visual cues or a small set of visual cues [2], we use a large set of automatically extracted visual nonverbal features as well as audio ones. Our experiments have been performed on a dataset where personality has not been studied before, where visual activity features provided the highest correlations and also the highest classification rates. For the classification of personality impressions, we use ridge regression. Ridge regression models provide a linear framework, which is able to handle large correlated feature spaces via the use of a regularization term, making it suitable for classification without a separate feature selection step. We further investigated the effect of the segment that is processed to predict personality. Although the external observers made their impressions based on a thin slice selected from the meeting, in the actual prediction task, the whole meeting is given, and one needs to decide on the part of the meeting to process. Whether thin slice impressions would generalize to the whole meeting for the prediction task is an open question.

In the next section, we present the data and the personality annotations used in our study. Section 3 explains the audio-visual nonverbal features. We discuss the correlations between the features and the personality impressions in Section 4. Section 5 presents the models that we use for personality prediction. We give the results of our experiments and conclude in Sections 6 and 7, respectively.

2. DATA AND ANNOTATIONS

2.1 Data

We used a subset from the Emergent LEAder (ELEA) corpus [12] for this study. The ELEA AV subset consists of audio-visual recordings of 27 meetings, in which the participants perform a winter survival task with no roles assigned. The participants in the task, as the survivors of an airplane crash, asked to rank 12 items to take with them to survive as a group. Participants first ranked the items individually, then as a group. The task itself promotes interactions among the participants in the group and each participant is encouraged to participate in the discussion and explain their preferences. While the corpus is originally designed to study leadership, the personality of each individual can be made evident through the discussion and negotiation parts of the interaction. There are 102 participants in total (six meetings with three participants and 21 meetings with four participants). Each meeting lasts around 15 minutes and recorded with two webcams and a microphone array. More details about the ELEA corpus can be found in [12, 13]. Figure 1 shows a snapshot from the ELEA AV corpus.

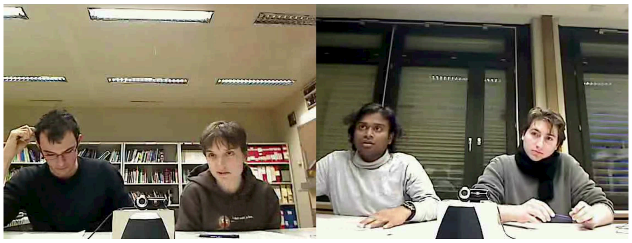


Figure 1: Snapshot from ELEA corpus

Figure 2: A snapshot from the cropped videos used in the annotations of the ELEA corpus

2.2 Personality Annotations

We have collected personality impressions of the external observers of the ELEA AV corpus. We used the Ten Item Personality Inventory (TIPI) for measuring the Big Five personality traits of the participants [14]. The TIPI questionnaire includes two questions per trait, answered on a 7-point Likert scale. For each participant, we have selected a one-minute segment from the meeting, which corresponds to the segment that includes the participant's longest turn. As we are interested in the personality of an individual within a meeting, we isolated the video of each participant: Only a single participant is visible. The audio on the other hand is intact and contains speech from all participants in that segment. Figure 2 shows a snapshot from the videos used in the annotations.

The video of each participant was observed and annotated by three different annotators. A total of five annotators annotated the whole dataset. The videos were shown muted and the annotators were given instructions to watch the videos without the audio. This was necessary as the meetings in the ELEA corpus were held in different languages (English, French and Spanish). Muting the audio helps to rule out any differences due to meeting language and annotators' language knowledge.

For each participant, the overall personality impression score for each trait is obtained by calculating the average of the scores of three annotators. The distribution of the average scores for each of the five personality traits is shown in Figure 3. The plots show that the impressions have distri-

Table 1: The mean and median values for the Big Five traits

	Mean	Median
Ext	4.06	3.92
Agr	4.68	4.67
Con	4.58	4.67
Emo	4.36	4.33
Ope	4.26	4.33

Table 2: Agreement between the annotators on the Big Five personality traits

	ICC(1,1)	ICC(1,k)	α
Ext	0.46	0.73	0.72
Agr	0.39	0.66	0.66
Con	0.07	0.19	0.19
Emo	0.02	0.05	0.09
Ope	0.28	0.54	0.54

butions of different characteristics. Table 1 shows the mean and median values for the Big Five traits. We observe that while for extraversion the distribution is more flat, the other four traits have a slight trend towards higher scores.

The agreement between the annotators for each trait is given in Table 2, in terms of Intra-Correlation Coefficient (ICC) and Cornbach's alpha (α) measures. The agreement is relatively higher for extraversion and moderate for agreeableness, and openness to experience. The results obtained here are parallel to the findings in the original study: as the TIPI questionnaire includes only two questions per trait, quantities such as alpha coefficients are known to be relatively low [14]. On the other hand, the very low agreement for the conscientiousness and emotional stability traits may relate to the relevance of the meeting domain to the investigated traits. The ELEA corpus depicts small group face-to-face conversations, in which not all the personality traits are necessarily encoded strongly [11], i.e. one could expect the extraversion trait to be strongly encoded and displayed in a conversation, whereas not so for conscientiousness.

3. NONVERBAL FEATURE EXTRACTION

Nonverbal behavioral cues are indicators of personality. For instance, in [15], it was shown that extraverted people use a louder voice, are more animated and expressive, use faster and more energetic gestures, and change their expression more. Many other papers in nonverbal communication suggest these findings [1].

Following this literature, we extracted audio and visual nonverbal features as descriptors of the nonverbal activity of participants. We used two approaches based on the segment of the meeting to be processed. First, we used the one-minute segment of each meeting shown to the annotators as the basis of our processing (Section 2.2). In the second approach, we used the whole meeting to extract the same list of features. The features obtained using the whole meeting are only used to report the results in Section 6.3. All other results are obtained using the features extracted based on the first approach (using one-minute segments). In either case, the features are extracted from each participant's activity in the given segment. The features are summarized in Table 3. In this section, we briefly present the audio and visual features that are used in the prediction model, and refer the reader to related papers for detailed information on audio-visual feature extraction.

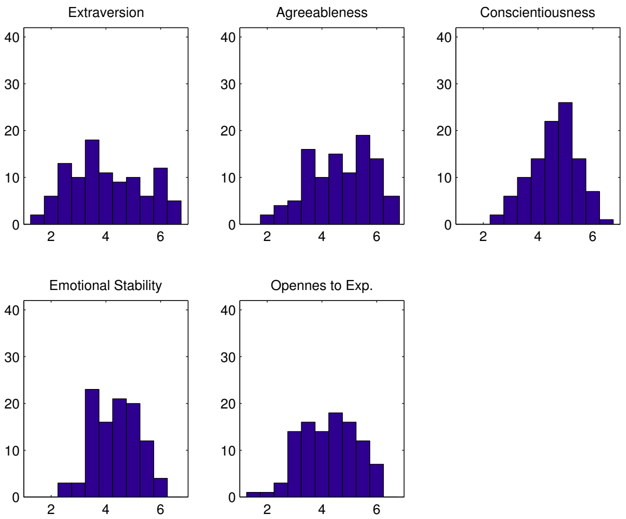


Figure 3: Histogram of the Big Five personality scores, annotated by external observers

3.1 Audio Features

Although the annotators watched the recordings without hearing the audio, we have extracted the audio features for each participant, because even without hearing the audio, one can understand some speaking related information, such as the speaking status, and how energetic a speaker can be. As discussed in Section 4, we have indeed observed correlations between several of the audio features and the personality trait impressions.

3.1.1 Speaking turn features

Speaking turn features are extracted from the binary segmentation that indicates the speaking status of each participant. This binary segmentation is provided by the microphone array that is used for the audio recordings, which performs speaker diarization and outputs the binary speech segmentation for each participant [12]. As speaking turn features, we have extracted Total Speaking Length (TSL), Total Speaking Turns (TST), Average Speaking Turn duration (AST). We have also extracted a filtered version of turns feature (TSTf), in which turns shorter than 2 seconds were not taken into account.

3.1.2 Prosodic features

Based on the binary speaker segmentation, we have obtained the speech signal for each participant. Overlapping speech segments were discarded, and only the segments with the participant being the sole speaker were considered for further processing. Two prosodic speech features, *energy* and *pitch*, were computed on the signal [12] and the following statistics were calculated for each feature: minimum, maximum, median, mean, standard deviation.

Table 3: Automatically extracted nonverbal features

Speaking Turn	TSL	Speaking Length
	TST	Speaking Turns
	AST	Average Speaking Turn duration
	TSTf	Speaking Turns (filtered)
Energy	Emin	Minimum Energy
	Emax	Maximum Energy
	Emed	Median Energy
	Emean	Mean Energy
	Estd	Energy Standard Deviation
Pitch	Pmin	Minimum Pitch
	Pmax	Maximum Pitch
	Pmed	Median Pitch
	Pmean	Mean Pitch
	Pstd	Pitch Standard Deviation
Visual Activity	THL	Head Activity Length
	THT	Head Activity Turns
	AHT	Average Head Activity turn duration
	TBL	Body Activity Length
	TBT	Body Activity Turns
	ABT	Average Body Activity turn duration
	stdHx	Std. Dev. head activity in x direction
	stdHy	Std. Dev. head activity in y direction
	stdB	Std. Dev. body activity
wMEI	wMEIent	wMEI entropy
	wMEImed	wMEI median
	wMEImean	wMEI mean
	wMEIqnt	wMEI quantile
	wMEImedZ	wMEI meanZ
	wMEImeanZ	wMEI meanZ
	wMEIqntZ	wMEI quantileZ
VFOA	AG	Attention Given
	AGspk	Attention Given while speaking
	AGlis	Attention Given while listening
	AR	Attention Received
	ARspk	Attention Received while speaking
	ARlis	Attention Received while listening
	VDR	Visual Dominance Ratio

3.2 Visual features

3.2.1 *Visual activity features*

Visual activity features characterize the bodily activity of the participant. We have used two different approaches to extract activity features. The first approach is based on head and body tracking and optical flow, which provides the binary head and body activity status and the amount of activity as well. Based on this information, we extracted features for head activity length (THL), head activity turns (THT), head activity average turn duration (AHT), body activity length (TBL), body activity turns (TBT), body activity average turn duration (ABT), and also the standard deviations of the head activity in x and y dimensions (stdHx, stdHy), and of the body activity (stdB).

3.2.2 *Motion template based features*

As a second approach, we have used weighted Motion Energy Images (wMEI) [12] as descriptors of spatio-temporal body activity and extracted several statistics as features, such as mean, median, 75% quantile, and entropy. Mean, median, and quantile statistics are also calculated by omitting zero intensity pixels in wMEI. We used the length of the meeting segment to normalize the images.

3.2.3 *Visual focus of attention features*

The visual focus of attention (VFOA) is estimated in [16], where a probabilistic framework was used to estimate the head location and pose jointly based on a standard state-space formulation. We have extracted the given and received attention for each participant based on the VFOA estimates, which indicate possible locations such as one of the other participants, the table area, or unfocused. Given attention is calculated as the amount of time the target participant is looking at the other participants. Received attention is calculated as the percentage of other participants looking at the target participant, averaged over the meeting duration. For given and received attention, we have also calculated two variants which take into account whether the target participant is speaking or not. As an additional feature, we calculated the Visual Dominance Ratio (VDR), which is defined as the ratio of looking while speaking over looking while listening [17]. As this feature uses audio and visual information, it is considered as a multimodal feature. The features extracted for each participant based on VFOA analysis can be summarized as follows: Attention given (AG); Attention given while speaking (AGspk); Attention given while listening (AGlis); Attention received (AR), Attention received while speaking (ARspk); Attention received while listening (ARlis), and the visual dominance ratio (VDR) (i.e. AGspk / AGsil). An important property of the VFOA based features is that they describe the behavior of the target participant in relation with the other participants in the meeting. With this property, these features describe the nonverbal interaction among the participants, whereas other nonverbal features presented above use only the audio or visual activity of the target participant.

4. CORRELATION ANALYSIS

We have calculated the correlation between the automatically extracted nonverbal features and the Big Five personality traits. Table 4 shows the Pearson correlation coefficients, with p values smaller than 0.05.

We see that even though the audio was muted to the annotators, there is still significant correlation between the audio features and the personality impression scores from the external annotators. This can be due to two reasons: First, the annotators are able to understand, to a large degree, whether the participant is speaking, due to lip movement. It can also be possible to understand whether the participant's speech is energetic by only looking at the visual channel. Second, there is inherent correlation between the audio and visual features. For example, a person speaking with high energy would have accompanying gestures and body movements. Regarding energy features, while minimum and maximum energy are negatively correlated with extraversion, mean energy is positively and more strongly correlated. Similar correlations are observed for openness to experience. Speaking turn features also have moderate positive correlation with extraversion, and negative correlation with agreeableness. Pitch does not seem to be correlated with any of the traits, except Pmin, which has low correlation with extraversion. This can be due to the muting of audio during the annotations, as unlike speaking turn or energy features, pitch is mostly identifiable in audio (with the exception of gender information). The correlations of audio features that we observe in our study are comparable with other studies, such as [18], except for the pitch features. We conclude that muting the audio during the annotations mostly affect the pitch features, among the audio features used here.

for visual features, highest correlations are observed between the visual features and extraversion. Body activity features, standard deviation of head and body, and wMEI based features are highly correlated with extraversion, and also with openness to experience (with less strength).

Among the VFOA based features, attention received while speaking has positive correlation with extraversion, while visual dominance ratio has negative correlation. This is an interesting result as it has been reported that dominant people have association to high visual dominance ratio values [17] and that dominance is also related to extraversion [11].

In general, significant correlations are observed between visual features and speaking energy features and extraversion and openness to experience trait impressions. Correlations are also observed for the agreeableness trait, for speaking turn features, and also for some of the VFOA based features. No correlation is observed for visual features for the agreeableness. Finally, the conscientiousness and emotional stability tratis essentially not captured by the features (but also keep in mind that their reliability is low).

5. INFERRING PERSONALITY

Ridge regression [19] estimates the regression coefficients using

$$\hat{\beta} = (X^T X + kI)^{-1} X^T y, \qquad (1)$$

where X is the feature matrix, k is the ridge parameter, I is the identity matrix, and y is the observation. The term "kI" acts as a regularizer with positive values of k. When k is 0, the model does not use any regularization term, thus is equal to the linear regression. The regularization term in the ridge regression, while introducing a bias, helps to reduce the variance of the estimate. It also helps to overcome the singularity problem in the case of linear regression, which occurs when the data is linearly dependent, i.e. features are highly correlated. With the introduction of the regularization term,$(X^T X + kI)$ is always invertible.

Table 4: Pearson correlation between features and Big Five personality traits

Features	Ext	Agr	Cons	Emo	Ope
TSL	0.30**	-0.30**			
TST					
TSTf	0.22*	-0.24*		0.21*	
AST	0.26*	-0.33***			
Emin	-0.20*	0.22*		-0.25*	
Emax	-0.22*				-0.24*
Emed					
Emean	0.39***				0.33***
Estd	0.24*				0.22*
Pmin	-0.20*				
Pmax					
Pmed					
Pmean					
Pstd					
THL					
AHT					
THT					
TBL	0.44***				0.22*
ABT			-0.28**		
TBT	0.47***				
stdHx	0.43***				
stdHy	0.37***				0.23*
stdB	0.37***				0.20*
wMEIent	0.40***				
wMEImed	0.26*				
wMEImean	0.44***				0.24*
wMEIqnt	0.49***				0.20*
wMEImedZ	0.27*				
wMEImeanZ	0.43***				0.25*
wMEIqntZ	0.48***				0.21*
AG					
AGspk					
AGlis	0.23*				
AR	0.27*				0.26*
ARspk	0.40***				0.22*
ARlis		0.24*			
VDR	-0.32***	0.27*			

***: p<<0.001, **: p<0.005, *:p<0.05

Ridge regression is a linear model and can be used as a classifier by using a threshold on the estimated values. In this case, the learned model is a linear classifier defined by a separating hyperplane. With this property, ridge regression classifier and a Support Vector Machine (SVM) with a linear kernel provide similar solutions and one would expect to have similar performance. From the practical aspects, in comparison to using SVMs, it is faster to train ridge regression models. In addition, we have observed that, on the ELEA data for personality prediction, ridge regression is less sensitive to the ridge parameter, thus easier to optimize, compared to optimizing the C parameter of linear SVM. Moreover ridge regression provides efficient calculations for leave-one-out cross validation, without the need of

Table 5: Regression results for extraversion trait. R^2 and MSE values for different feature sets are presented. The highest R^2 is shown in bold; bold and italic indicate $R^2 > 0.2$.

Ext	ST	E	P	VA	wMEI	VFOA	All
R^2	0.12	0.09	0.03	***0.27***	**0.31**	0.13	***0.30***
MSE	1.62	1.69	1.80	***1.36***	**1.28**	1.61	***1.29***

re-training, making it even easier to optimize the parameters [19].

6. EXPERIMENTS AND RESULTS

We study personality prediction from two angles. First, we look at both regression and classification problems and assess the performance of prediction for the three of the Big Five personality traits (extraversion, agreeableness, openness to experience) on the annotated one-minute segments. We excluded the conscientiousness and emotional stability traits due to low inter-annotator agreement and low correlation with features. In the regression experiments, we assess the models on predicting the actual personality impression scores for each trait. For the classification experiments, we convert the problem to a classification problem by defining two classes for each trait as high and low, based on the scores. Second, we evaluate and assess the performance of the same models when generalized to the whole meeting instead of a short segment, i.e. when the whole meeting is given as the segment to process.

In the experiments below, we use leave-one-out cross validation and report the average accuracy over all folds. We normalize the data such that each feature has zero mean and one standard deviation. The ridge parameter is optimized using a nested cross validation scheme, with values in the range of [2, 150]. The parameters of SVM are optimized similarly with a nested cross validation scheme, with C parameter values selected from the $[2^{-3}, 2^3]$ range. We use different feature sets as summarized in Table 3, namely Speaking Turn (ST), Energy (E), Pitch (P), Visual Activity (VA), wMEI, VFOA features, and the combination of all features (ALL). These acronyms for feature sets are also used in Tables 5 - 8 and in Figures 4(a) - 4(b).

6.1 Regression

For the regression experiments, we have used the mean personality impression scores, range between 1 to 7, as calculated from the annotations. The R^2 and Mean Squared Error (MSE) values are given in Table 5. We only show the results for the extraversion trait. For the other two traits, the R^2 values are less than 0.1. The bold value indicates the feature set with the highest R^2, and bold and italic values indicates $R^2 > 0.2$ .We obtain R^2 values as high as 0.31 with the wMEI features. Visual features provide a higher R^2 than the audio ones. Combining all features does not outperform the use of the best single cues.

6.2 Classification

Table 6: Classification results with ridge regression classifier using 0/1 labels as scores. For each trait, the accuracies that are significantly different than the baseline are shown in bold and italic. The highest accuracy is shown in bold.

R_{LBL}	ST	E	P	VA	wMEI	VFOA	All
Ext	57.8	***65.7***	60.0	***71.2***	***68.6***	53.9	**74.5**
Agr	50.0	57.8	51.0	54.9	50.0	46.1	52.0
Ope	52.0	52.9	52.9	52.9	58.8	55.9	47.1

Table 7: Classification results with ridge regression classifier using original personality impression scores. For each trait, the accuracies that are significantly different than the baseline are shown in bold and italic. The highest accuracy is shown in bold.

R_{SCR}	ST	E	P	VA	wMEI	VFOA	All
Ext	53.9	57.8	54.9	***71.6***	***66.7***	55.9	**72.6**
Agr	49.0	48.0	59.8	52.0	51.0	53.9	57.8
Ope	49.0	48.0	47.1	52.0	51.0	54.9	52.9

For the classification experiments, we have used the median of the scores as the cutoff point and set label 1/0 for scores greater/smaller than the median score, e.g. to represent people scoring high/low in extraversion, respectively. This procedure creates balanced classes for the two labels and a random baseline accuracy of 50%. The same procedure is applied for each of the personality traits. In Tables 6 - 8 the bold values indicate the highest accuracy and both the italic and bold values indicate the results that are statistically significant over the baseline.

For training a ridge regression classifier, we used two methods. First, we used 0/1 labels as the scores and trained a regression model (R_{LBL}). For prediction, the trained model is used to make an estimate and the predicted label is set to 0 if the estimated score is less than 0.5, otherwise the predicted label is set to 1. The second method uses the original personality impression scores when training the model and the median score is used as the threshold for prediction (R_{SCR}).

The classification results with R_{LBL} for each of the personality traits are given in Table 6. Only accuracies above 62.7% are considered significantly different (with 99% confidence level) than the 50% baseline. For extraversion, we obtain accuracies as high as 74.5%. For the other traits, the results are not significantly different than the random baseline.

In Table 7, we show the classification results with R_{SCR}. The results are slightly lower than that of R_{LBL} but the difference is not statistically significant for the best results. Comparing the classification results with the R^2 and MSE values obtained in regression task (see Table 5), we see that a higher R^2 does not mean a higher accuracy in classification. Although wMEI has the highest R^2 value, its classification accuracy is lower than VA and ALL.

For comparison purposes, we show the classification results with SVMs with linear kernel in Table 8. We have also performed the experiments with SVMs with Gaussian kernel, however the results are similar to SVMs with linear kernel, thus we only report the results with linear kernel. The results are slightly lower than the ridge regression results but the difference is not statistically significant. The main advantage of using ridge regression models for classification

Table 8: Classification results using SVM. For each trait, the accuracies that are significantly different than the baseline are shown in bold and italic. The highest accuracy is shown in bold.

	ST	E	P	VA	wMEI	VFOA	All
Ext	52.9	***67.7***	55.9	***70.6***	***64.7***	52.0	**71.6**
Agr	53.9	55.9	44.1	56.9	54.9	55.9	49.0
Ope	41.2	43.1	51.0	47.1	45.1	32.4	49.0

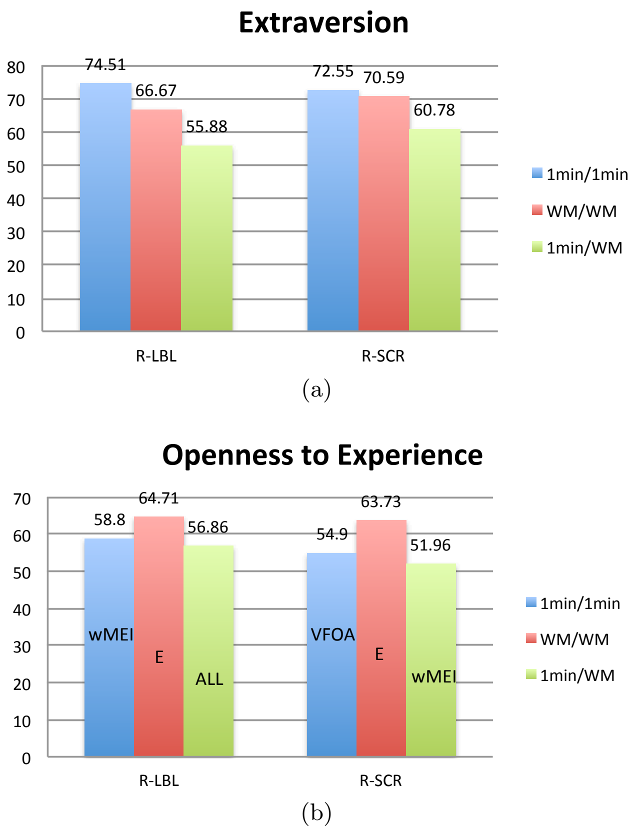
is that the ridge regression better handles correlated feature spaces via the regularization term. In our case, most features are highly correlated with each other and SVMs would perform better if a feature selection step prior to classification was applied, which results in better feature utilization. Given the moderate size of our dataset, feature selection methods generally result in overfitting. Instead, we perform ridge regression without feature selection and our results show that the models are able to generalize relatively well.

6.3 From slices to whole meetings

The personality impressions from external observers were obtained on a one-minute segment extracted from each meeting, as discussed in Section 2.2. An open question is to what extent these personality impressions apply to participant's behavior in longer duration. As personality is considered to be a stable trait of a person, are external observers assign attributes that are only reflective of the segment that they watch, or do they generalize? We explore this problem in this study and compare three different setups:

- 1min/1min: Only the one-minute segment that the annotators watched to make their personality impressions are used to extract features and to train the models. Any other part of the meeting is discarded and not used in any stage of the modeling. At the test phase, for each meeting in the test data, the same one-minute segment that the annotators watched to make their personality impressions is used to extract the features for the test examples. In this setup, we use the same information that the annotators used to make their judgments. The experiments in the previous section uses this setup.
- WM/WM: Instead of using the one-minute segment, we use the whole meeting to extract the features. Both training and test examples use the features extracted form the whole meeting. In this setting, we use extra information that the annotator did not have any access while making their judgments, both in training and test phases.
- 1min/WM: In this setup, the training examples have features extracted from the one-minute segment, while the test examples have features extracted from the whole meeting.

Using the above setups, we investigate the generalization ability of our models when the unit of processing has changed. We have used both R_{LBL} and R_{SRC} classifiers and run for each setup, for each personality trait. We obtained accuracies that are significantly higher than the random baseline only for the extraversion and openness to experience traits. Figures 4(a) and 4(b) show the accuracies for extraversion and openness to experience traits, for each model. The feature group that gives the reported accuracy is also shown. We see that, for extraversion, there is a decrease in the accuracy if the unit of processing is different than the annotated segment. The accuracy of WM/WM setup is only slightly lower than 1min/1min, indicating that the extraversion impressions generalize to longer segments. For extraversion, visual features (VA and wMEI) provide higher accuracies than audio features (see also Tables 6-8). For the openness to experience trait, the WM/WM setup using energy features provide the highest accuracies. While the accuracy of 1min/1min setup is not significantly higher than the random baseline, WM/WM accuracy is significant. One possible explanation is that the set of features we used in this study do not include the specific cues that the annotators used to make their impressions for openness to experience on the given one-minute segment. Despite this fact, the energy features provide reliable information when measured on a longer duration. For both traits, using the same unit of processing in training and test instances (1min/1min and WM/WM) achieve higher accuracies than using different units (1min/WM).



(a)

(b)

Figure 4: Generalization performance of the models, from impressions on one-minute segments to whole meeting for (a) extraversion and (b) openness to experience traits.

7. CONCLUSIONS

We have presented a systematic analysis on automatic prediction of personality impressions in small group meetings where teams play a winter survival task. Our findings indicate that extraversion is the best predicted trait in this domain, which is a well known fact supported both by social psychology literature [11] and also by other works on automatic personality impression prediction [5, 3, 4]. Visual features, particularly visual activity features, better represent extraversion impressions and achieve better performance both in regression and classification tasks in comparison to audio features. We have also observed that among the audio features, energy features have higher accuracies. We have also investigated and discussed the effect of the unit of processing on the personality prediction performance. During annotations, external observers make their personality impressions based on a thin sliced segment from a meeting. We investigated whether the personality impressions of the annotators, judged on a thin-slice, generalize to a larger segment from which the features are extracted in an automatic prediction setting. Our results show that for the extraversion trait, predictions based on the whole meeting can be made, with a slightly lower accuracy. For the openness to experience trait, using the whole meeting provides even higher accuracies. As a future dimension of this study, the concept of generalization can be further investigated by using different representations of a meeting, for instance as a collection of multiple segments.

Acknowledgments

O. Aran thanks the support of the Swiss National Science Foundation (SNSF) Ambizione fellowship through the SOBE project (PZ00P2_136811). D. Gatica-Perez thanks the support of the SNSF through the SONVB project. We thank D. Sanchez-Cortes and D. Jayagopi (Idiap) for previous work on data collection and feature extraction.

8. REFERENCES

[1] M. L. Knapp and J. A. Hall, *Nonverbal Communication in Human Interaction*. Wadsworth, Cengage Learning, 2008.

[2] F. Pianesi, N. Mana, and A. Cappelletti, "Multimodal recognition of personality traits in social interactions," in *ICMI*, 2008.

[3] B. Lepri, S. Ramanathan, K. Kalimeri, J. Staiano, F. Pianesi, and N. Sebe, "Connecting meeting behavior with extraversion - a systematic study," *T. Affective Computing*, vol. 3, no. 4, pp. 443–455, 2012.

[4] J. Staiano, B. Lepri, S. Ramanathan, N. Sebe, and F. Pianesi, "Automatic modeling of personality states in small group interactions," in *ACM Multimedia*, 2011, pp. 989–992.

[5] J.-I. Biel and D. Gatica-Perez, "The youtube lens: Crowdsourced personality impressions and audiovisual analysis of vlogs," *IEEE Transactions on Multimedia*, 2012.

[6] J.-I. Biel, L. Teijeiro-Mosquera, and D. Gatica-Perez, "Facetube: predicting personality from facial expressions of emotion in online conversational video," in *Proceedings International Conference on Multimodal Interfaces (ICMI-MLMI)*, 2012.

[7] G. Mohammadi, A. Origlia, M. Filippone, and A. Vinciarelli, "From speech to personality: mapping voice quality and intonation into personality differences," in *ACM Multimedia*, 2012, pp. 789–792.

[8] L. M. Batrinca, N. Mana, B. Lepri, F. Pianesi, and N. Sebe, "Please, tell me about yourself: automatic personality assessment using short self-presentations," in *ICMI*, 2011, pp. 255–262.

[9] G. Chittaranjan, J. Blom, and D. Gatica-Perez, "Mining large-scale smartphone data for personality studies," *Personal and Ubiquitous Computing*, vol. 17, no. 3, pp. 433–450, 2013.

[10] D. O. OlguÃŇn, P. A. Gloor, and A. Pentland, "Capturing individual and group behavior with wearable sensors." in *AAAI Spring Symposium: Human Behavior Modeling*. AAAI, 2009, pp. 68–74.

[11] R. Gifford, *The SAGE Handbook of Nonverbal Communication*. SAGE Publications, Inc., 2006, ch. Personality and Nonverbal Behavior: A Complex Conundrum, pp. 159–181.

[12] D. Sanchez-Cortes, O. Aran, M. Mast, and D. Gatica-Perez, "A nonverbal behavior approach to identify emergent leaders in small groups," *Multimedia, IEEE Transactions on*, vol. 14, no. 3, pp. 816–832, 2012.

[13] D. Sanchez-Cortes, O. Aran, and D. Gatica-Perez, "An audio visual corpus for emergent leader analysis," in *ICM-MLMI'11: Workshop on Multimodal Corpora for Machine Learning: Taking Stock and Road mapping the Future*, Nov 2011.

[14] S. D. Gosling, P. J. Rentfrow, and W. B. Swann, "A very brief measure of the big-five personality domains," *Journal of Research in Personality*, vol. 37, pp. 504–528, 2003.

[15] R. Lippa, "The nonverbal display and judgment of extraversion, masculinity, femininity, and gender diagnosticity: A lens model analysis," *Journal of Research in Personality*, vol. 32, no. 1, pp. 80 – 107, 1998.

[16] D. Sanchez-Cortes, O. Aran, D. B. Jayagopi, M. Schmid Mast, and D. Gatica-Perez, "Emergent leaders through looking and speaking: from audio-visual data to multimodal recognition," *Journal on Multimodal User Interfaces*, 2012.

[17] J. Dovidio and S. Ellyson, "Decoding visual dominance: Attributions of power based on relative percentages of looking while speaking and looking while listening," *Social Psychology Quarterly*, vol. 45, no. 2, pp. 106–113, 1982.

[18] J.-I. Biel, O. Aran, and D. Gatica-Perez, "You are known by how you vlog: Personality impressions and nonverbal behavior in youtube," in *Proceedings of AAAI International Conference on Weblogs and Social Media (ICWSM)*, 2011.

[19] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, ser. Springer Series in Statistics. New York, NY, USA: Springer New York Inc., 2001.