

Prediction of the Leadership Style of an Emergent Leader Using Audio and Visual Nonverbal Features

Supplementary Material

Cigdem Beyan, *Member, IEEE*, Francesca Capozzi, Cristina Becchio, and Vittorio Murino, *Senior Member, IEEE*

This supplementary material contains more information regarding the dataset [1] utilized, and the nonverbal features (NFs) used. It also reports the results obtained from the correlation analysis between the extracted NFs and the results of SYMLOG (the SYstematic method for the Multiple Level Observation of Groups) questionnaire (friendliness sub-scale).

I. DATASET

Each meeting contains a collective discussion such that each member was invited to contribute e.g., participants are acting similarly, nothing in the task is leading to one behaving different than the others. Participants are sitting during the whole group task and there is no restriction given to any participant regarding how they should/should not behave, i.e. meetings are completely in-the-wild. In several meetings, it is possible to observe that participants can speak at the same time and there are many speaking interruptions as well.

As mentioned in the main manuscript, participants in each meeting performed winter or desert survival [2]. Regarding survival task, it is possible to evaluate the performance of participants as a group. In detail, the participants had to rank 12 items in order of importance for their survival in a dramatic scenario (for more details, see [2]). Their ranking can be compared to an optimal ranking which made by experts. The smallest the difference between the two rankings (participants and experts), the better the performance. Therefore, the optimal performance score is zero. The random performance can be calculated as 12×6 which is the average performance. The comparison between participants' performance and the optimal and the random performances allows us to evaluate the performance of participants. Therefore, we compared the results of each meeting both to the optimal performance and to a random performance by applying one-sample, two-tailed t-test. The results showed that meetings have an average performance that is lower than the optimal ($p < 0.0001$) and have an average performance that is higher than the random

performance ($p < 0.0001$). Overall, this suggests that the surveillance tasks applied were difficult enough to engage participants in the interaction, but they were not too hard such that allow participants doing it properly.

II. NONVERBAL FEATURES (NFs)

A visualization of the dataset in terms of nonverbal features and the classes (democratic leadership style (LS), autocratic LS and not-a-leader) are given in Figure 1a and 1b for the meeting segments and the holistic data, respectively. In this figure, all audio and video based NFs (in total 54) were used since the classification results when they used all together were found the best as compared to using any other feature combinations. t-SNE [3] were utilized to represent the features in two-dimensions.

As seen in Figure 1, for both type of data, the classes are overlapping in the two-dimensional feature space which means that the classification tasks are not very simple. Therefore, complex classification methods, such as multiple kernel learning [4], [5] as applied in this study, are beneficial to obtain accurate classification results.

III. CORRELATION ANALYSIS

In Table I, the Pearson's Correlation Coefficients [6] (shown as Corr-All and Corr-Lea) with significance (shown as Sig-All and Sig-Lea) between results derived from SYMLOG friendliness and nonverbal features using the data belong to all participants (shown as All) and belong to only leaders (shown as Lea) are given. In this table, *H* means high correlation (correlation coefficients from 0.5 to 1.0 or -0.5 to -1.0), *M* represents medium correlation (correlation coefficients from 0.3 to 0.5 or -0.3 to -0.5), *L* presents low correlation (correlation coefficients from 0.1 to 0.3 or -0.1 to -0.3), *N* refers to no correlation, *S* is used if there is a significance (p -value < 0.05) and *NS* is used if there is no significance (p -value ≥ 0.05). Since, SYMLOG results correspond to holistic meetings only, the evaluation was performed using holistic data.

The results showed that many nonverbal features were not correlated (18 out of 68 nonverbal features when data of all participants used and 9 out of 68 when data of leaders used) or low correlated but without significance (39 and 29 out of 68 using all participants' data and leaders only, respectively). 11 and 10 nonverbal features out of 68 were detected significantly

C. Beyan and V. Murino are with Pattern Analysis and Computer Vision (PAVIS), Istituto Italiano di Tecnologia (IIT), Genoa, 16163, Italy. V. Murino is also with Department of Computer Science, Univeristy of Verona, Verona, Italy (e-mail: cigdem.beyan@iit.it; vittorio.murino@iit.it).

F. Capozzi is with Department of Psychology, in McGill University, Montreal, QC, 1205, Canada (e-mail: francesca.capozzi@mcgill.ca).

C. Becchio is with Robotics, Brain and Cognitive Sciences (RBCS), Istituto Italiano di Tecnologia (IIT), Genoa, 16163, Italy and Department of Psychology in University of Turin, Torino, 10124, Italy (e-mail: cristina.becchio@iit.it).

TABLE I: Correlation and significance results between SYMLOG friendliness and nonverbal features. Significantly correlated nonverbal features are emphasized in bold-face. (See text for more details.)

Nonverbal Features	Corr-All	Sig-All	Corr-Lea	Sig-Lea
TW	-0.07 (N)	NS	-0.10 (L)	NS
TME	-0.09 (N)	NS	-0.11 (L)	NS
TWwNoME	-0.05 (N)	NS	-0.08 (N)	NS
TL	0.10 (L)	NS	0.43 (M)	NS
TInitME	0.02 (N)	NS	0.19 (L)	NS
TIntCME	0.13 (L)	NS	0.38 (M)	NS
TLNoME	0.15 (L)	NS	0.47 (M)	NS
RTWTL	-0.10 (L)	NS	-0.27 (L)	NS
THL	-0.08 (N)	NS	-0.23 (L)	NS
THT	-0.08 (N)	NS	-0.34 (M)	NS
AHT	0.26 (L)	NS	0.28 (L)	NS
stdHx	0.11 (L)	NS	0.35 (M)	NS
stdHy	-0.04 (N)	NS	0.29 (L)	NS
TBL	-0.05 (N)	NS	-0.07 (N)	NS
TBT	-0.03 (N)	NS	0.07 (N)	NS
ABT	0.28 (L)	NS	0.20 (L)	NS
stdB	0.02 (N)	NS	-0.10 (L)	NS
TMSL	-0.22 (L)	NS	-0.38 (M)	NS
TSSL	-0.11 (L)	NS	-0.18 (L)	NS
FMSSL	-0.05 (N)	NS	-0.21 (L)	NS
TST	-0.16 (L)	NS	-0.32 (M)	NS
ASTD	0.11 (L)	NS	0.11 (L)	NS
TSI	-0.42 (M)	S	-0.46 (M)	NS
TUI	0.10 (L)	NS	0.39 (M)	NS
BSI	-0.28 (L)	NS	-0.44 (M)	NS
BUI	-0.18 (L)	NS	-0.02 (N)	NS
TSA	-0.20 (L)	NS	-0.21 (L)	NS
TSBI	0.05 (N)	NS	0.14 (L)	NS
SFG	-0.40 (M)	S	0.25 (L)	NS
FTSS	-0.15 (L)	NS	-0.30 (M)	NS
RSIT	-0.36 (M)	S	-0.39 (M)	NS
RUIT	0.09 (N)	NS	0.38 (M)	NS
TE	-0.53 (H)	S	-0.74 (H)	S
MinE	-0.22 (L)	NS	-0.49 (M)	NS
MaxE	-0.08 (N)	NS	-0.51 (H)	NS
MedE	-0.60 (H)	S	-0.81 (H)	S
MeanE	-0.58 (H)	S	-0.81 (H)	S
StdE	-0.49 (M)	S	-0.76 (H)	S
TENoOver	-0.51 (H)	S	-0.72 (H)	S
MinENoOver	-0.23 (L)	NS	-0.52 (H)	NS
MaxENoOver	-0.10 (N)	NS	-0.49 (M)	NS
MedENoOver	-0.60 (H)	S	-0.82 (H)	S
MeanENoOver	-0.59 (H)	S	-0.81 (H)	S
StdENoOver	-0.49 (M)	S	-0.75 (H)	S
MinP	-0.19 (L)	NS	-0.63 (H)	S
MaxP	0.03 (N)	NS	NaN (N)	NS
MedP	-0.20 (L)	NS	-0.15 (L)	NS
MeanP	-0.20 (L)	NS	-0.14 (L)	NS
StdP	0.13 (L)	NS	0.22 (M)	NS
MinPNoOver	-0.18 (L)	NS	-0.63 (H)	S
MaxPNoOver	0.06 (N)	NS	NaN (N)	NS
MedPNoOver	-0.21 (L)	NS	-0.15 (L)	NS
MeanPNoOver	-0.20 (L)	NS	-0.14 (L)	NS
StdPNoOver	0.14 (L)	NS	0.22 (L)	NS
TLwS	-0.15 (L)	NS	-0.25 (L)	NS
TLwNS	0.19 (L)	NS	0.55 (H)	NS
TWwS	-0.12 (L)	NS	-0.18 (L)	NS
T2WwS	-0.16 (L)	NS	-0.23 (L)	NS
T3WwS	-0.03 (N)	NS	0.06 (N)	NS
RTLwSTLwSS	-0.10 (L)	NS	-0.19 (L)	NS
RTLwSTLsNS	-0.19 (L)	NS	-0.41 (M)	NS
TMEwS	-0.17 (L)	NS	-0.19 (L)	NS
TVAwS	-0.18 (L)	NS	-0.22 (L)	NS
THATwS	-0.10 (L)	NS	-0.04 (N)	NS
THTwS	-0.23 (L)	NS	-0.15 (L)	NS
AVATwS	0.21 (L)	NS	0.22 (L)	NS
TVAIwS	-0.19 (L)	NS	-0.28 (L)	NS
TVAFGwS	-0.21 (L)	NS	0.09 (N)	NS

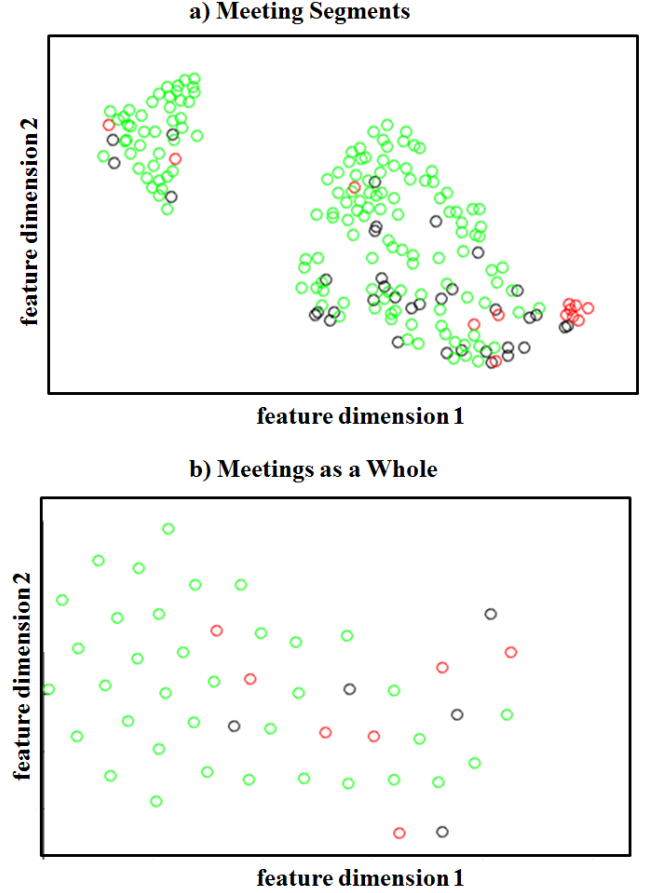


Fig. 1: The visualization of the audio-visual data: a) meeting segments, b) holistic meetings. Green, red and black circles show the data points belong to not-a-leader, autocratic LS and democratic LS, respectively.

medium or high correlated with SYMLOG friendliness results when the data of all participants and only leaders were used, respectively. Additionally, 17 features were identified as medium correlated without significance while 3 features were detected as high correlated without significance when only leaders data were analyzed. The medium and high correlation without significance means that the amount of data used was not enough to make a conclusion regarding that feature's correlation which shows that more samples are needed for their analysis.

The nonverbal features found as significantly correlated (medium and high) were TSI, SFG, RSIT, TE, MedE, MeanE, StdE, TENoOver, MedENoOver, MeanENoOver, StdENoOver, MinP, and MinPNoOver (see the main manuscript for their definitions) which are all audio-based, i.e. belong to speaking activity or prosodic features.

REFERENCES

- [1] C. Beyan, N. Carissimi, F. Capozzi, S. Vascon, M. Bustreo, A. Pierro, C. Becchio, and V. Murino, "Detecting emergent leader in a meeting environment using nonverbal visual features only." *ACM ICMI*, 2016, pp. 317–324.
- [2] D. Johnson and F. Johnson, *Joining together: Group theory and group skills*. Prentice-Hall, Inc., 1991.

- [3] V. der Maaten Laurens and G. Hinton, "Visualizing data using t-sne." *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.
- [4] M. Gonen and E. Alpaydin, "Localized multiple kernel learning," in *ICML*, 2008, pp. 352–359.
- [5] M. Gonen and E. Alpaydin, "Multiple kernel learning algorithms," *Journal of Machine Learning Research*, vol. 12, pp. 2211–2268, 2011.
- [6] K. Pearson, "Notes on regression and inheritance in the case of two parents." *Proceedings of the Royal Society of London*, 1895, pp. 240–242.