Emergent Leadership Detection Across Datasets

Philipp Müller
Max Planck Institute for Informatics
Saarland Informatics Campus
pmueller@mpi-inf.mpg.de

Andreas Bulling
University of Stuttgart
Institute for Visualisation and Interactive Systems
andreas.bulling@vis.uni-stuttgart.de

ABSTRACT

Automatic detection of emergent leaders in small groups from nonverbal behaviour is a growing research topic in social signal processing but existing methods were evaluated on single datasets – an unrealistic assumption for real-world applications in which systems are required to also work in settings unseen at training time. It therefore remains unclear whether current methods for emergent leadership detection generalise to similar but new settings and to which extent. To overcome this limitation, we are the first to study a cross-dataset evaluation setting for the emergent leadership detection task. We provide evaluations for within- and cross-dataset prediction using two current datasets (PAVIS and MPIIGroupInteraction), as well as an investigation on the robustness of commonly used feature channels and online prediction in the cross-dataset setting. Our evaluations show that using pose and eye contact based features, cross-dataset prediction is possible with an accuracy of 0.68, as such providing another important piece of the puzzle towards real-world emergent leadership detection.

CCS CONCEPTS

• **Applied computing** → **Psychology**.

KEYWORDS

social signal processing, emergent leadership detection

ACM Reference Format:
Philipp Müller and Andreas Bulling. 2019. Emergent Leadership Detection Across Datasets. In *2019 International Conference on Multimodal Interaction (ICMI '19), October 14-18, 2019, Suzhou, China.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3340555.3353721

ICMI '19, October 14-18, 2019, Suzhou, China

ACM ISBN 978-1-4503-6860-5/19/10…$15.00
https://doi.org/10.1145/3340555.3353721

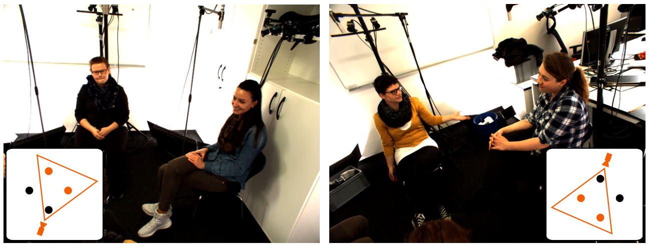

Figure 1: Illustration of the recording setup of the MPI-IGroupInteraction dataset [18]. The selected view and corresponding visible participants are shown in orange.

1 INTRODUCTION

Emergent leaders are group members who naturally obtain a leadership position through interaction with the group, and not via a higher authority [24]. Even without formal authority, emergent leaders are important for group performance [11, 16], and as a result automatic identification of emergent leaders in group interactions is potentially beneficial in organisational research, in the context of assessment centres [14], or for robots and intelligent agents that are supposed to interact with a group naturally. Consequently, the detection of emergent leaders is a growing topic in social signal processing [6, 12, 23]. These studies used nonverbal behaviour to detect emergent leaders in group interactions, which is supported by a large body of work connecting emergent leadership and nonverbal behaviour [1, 13, 15].

While existent methods on emergent leadership detection in small groups showed reasonable performance, they all make the assumption that training and testing data come from the same distribution. This assumption is unrealistic for application scenarios in which a system is required to detect emergent leaders in slightly different social situations for which no labelled data is available. Until now, it remains unclear whether such cross-dataset leadership detection is possible with sufficient accuracy.

Specifically, emergent leadership detection in small groups of unaugmented people has only been investigated separately on two datasets employing very similar tasks, effectively ignoring the crucial cross-dataset setting. The ELEA dataset [23] consists of meetings of three or four people each, in which participants are instructed to come up with a joint solution for the winter survival task. Work on ELEA investigated emergent leadership detection from recordings of the meetings, by using audio- and visual or multi-modal features [22, 23], and more recently by using features obtained from a co-occurrence mining procedure [20]. Kindiroglu et al. investigated domain adaptation and multi-task learning for leadership- and extraversion prediction on ELEA using video blogs with personality annotations [17]. Their work is different to the cross-dataset setting described above, as they assumed access to leadership ground truth on ELEA.

The PAVIS dataset [6] consists of groups of four people each either performing a winter- or a desert survival task. Research on the dataset focussed on detecting emergent leaders from nonverbal features only [6], using multiple kernel learning [4], or using body pose based features [7]. Further studies improved emergent leadership detection on PAVIS by using deep visual activity features [9], or by employing sequential analysis [8]. In addition, the dataset has been used to predict the leadership style of emergent leaders [5, 9].

Recently, the MPIIGroupInteraction dataset was recorded to study low rapport detection in small groups [18]. Although emergent leadership was rated, no corresponding detection approach was proposed. This dataset is particularly interesting for emergent leadership detection, as opposed to the rather constrained tasks on ELEA and PAVIS, participants engaged in open-ended discussions.

In this paper, we move one step closer to an emergent leadership detection system that can be applied in novel social situations without additional labelling effort. We investigate emergent leadership detection across situations using two recent datasets [6, 18] both featuring small group interactions but differing in participants' tasks, language, and nationality. Our specific contributions are twofold: We are the first to study emergent leadership detection in a cross-dataset setting, thereby achieving state-of-the-art results on MPIIGroupInteraction [18]. Furthermore, we conduct extensive evaluations providing insights into the usefulness of different features and the feasibility of an online prediction system.

2 DATASETS

To study cross-dataset emergent leadership detection, we utilise the PAVIS [6] and the MPIIGroupInteraction [18] datasets of small group interactions. We could not include ELEA because we found inconsistencies in the mapping between ground truth and videos that could not be resolved with the authors before submission.

PAVIS

The PAVIS dataset [6] consists of 16 interactions of four Italian speaking unacquainted participants each. Each group performed either a winter- or a desert survival task, in which participants had to agree on a ranking of the usefulness of items in a survival situation. Each participant was recorded by a frontal-facing camera and a lapel microphone. Interactions lasted from 12 to 30 minutes, resulting in a total corpus length of 393 minutes. All recordings were divided into segments of four to six minutes and subsequently annotated for emergent leadership. In line with previous work [9], we exclude four recordings due to audio problems, resulting in 12 meetings and 48 participants. We use PAVIS as a source dataset, as the segment-based annotation yields more training data than is available on MPIIGroupInteraction [18].

MPIIGroupInteraction

MPIIGroupInteraction consists of 22 group interactions in German, each consisting of three- to four unacquainted participants. In contrast to the rather constrained winter- or desert survival task on the PAVIS dataset [6], participants had an open-ended discussion. The meetings were recorded by eight frame-synchronised cameras, two of them placed behind every participants in order to cover all other participants in their field of view (see Figure 1). To record audio, one microphone was placed in front and slightly above participants' heads. Each group was discussing for roughly 20 minutes, resulting in more than 440 minutes of audio-visual recordings in total. After the interaction, each participant rated every other participant on a leadership scale ("PLead" as in [23]). We use the aggregate ratings for each participant to identify the ground truth emergent leader.

3 METHOD

To detect emergent leaders, we use Support Vector Machines and nonverbal features from gaze, body pose, face and speaking activity. We give a concise description of the method here and refer to the supplementary material for further details.

Nonverbal Feature Extraction

VFOA Features. To compute features based on the visual focus of attention (VFOA), we first perform eye contact detection, i.e. detecting at which other persons' face a target person is looking at a given moment in time. To this end, we employ the recently introduced method by Müller et al. [19], which performs unsupervised eye contact detection in small group interactions by exploiting natural conversational gaze behaviour in a weak labelling step. Based on these eye contact detections, we extract 15 VFOA features as described in [6]. While the features we compute on top of eye contact detections are the same as in [6], in the work of Beyan et al. they are based on VFOA detections using head pose.

Body Pose Features. We estimate body poses of participants using OpenPose [10] and follow the approach taken in [7] for pose feature computation. This approach yields a 80-dimensional featureset consisting of statistical measures based on the angles between detected body joints.

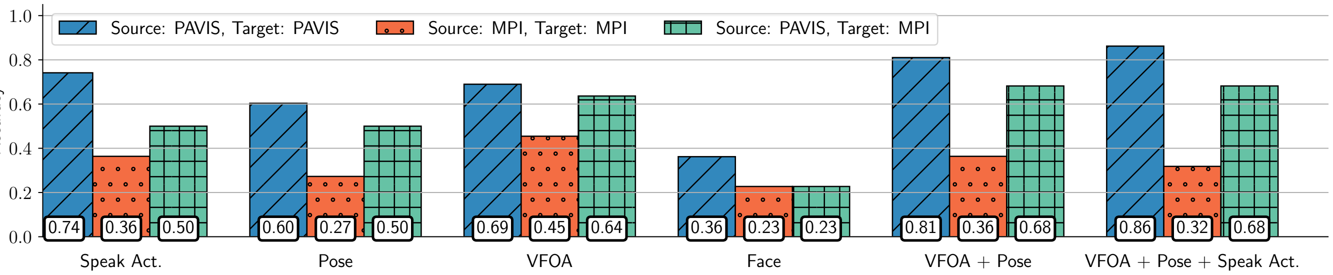

Figure 2: Performance of different featuresets when either training and testing on the same dataset, or training on PAVIS and testing on MPIIGroupInteraction. Random baseline for PAVIS as target is 0.25, for MPIIGroupInteraction as target 0.29.

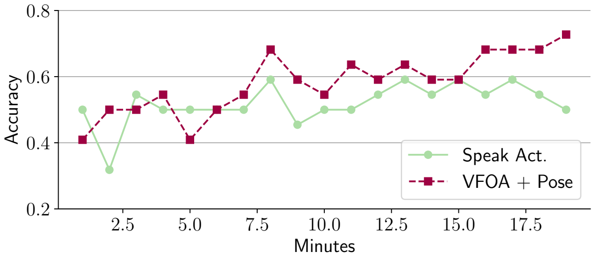

Figure 3: Performance of different featuresets when training on PAVIS and testing on MPIIGroupInteraction, depending on the size of the time window that is used for analysis (starting from the beginning). Random baseline is at 0.29.

Facial Features. We use OpenFace [2, 3] to extract facial action units (AUs) and subsequently follow the approach described in [18] for low rapport detection. We specifically extract the means of AU activations and intensities and the mean and standard deviation of a "facial positivity indicator".

Speaking Activity Features. To evaluate the importance of speaking activity, we implement features used in previous work [22], which encode the total speaking time of a participant, the number of speaking turns of a participant, the total number of times a participant interrupts other participants, and the average duration of a participants' speaking turns.

Classification

In line with previous work [7, 18], we use Support Vector Machines (SVMs) with radial basis function (RBF) kernels. To obtain a single predicted leader for each interaction during test time, we obtain probability estimates using Platt scaling [21] and select the participant with the highest probability as the predicted emergent leader. We choose the regularisation parameter C of the SVM via cross-validation on the source dataset (PAVIS), and set the parameter γ of the rbf kernel to the default value $1/n_{feats}$.

While normalising the training data by subtracting the mean and dividing by the standard deviation computed on the whole source dataset, we normalise each test interaction in the target dataset separately. In preliminary experiments, this way of normalising data has proven to be crucial. We refer to the supplementary material for a detailed discussion.

When employing several featuresets for classification, we always use late fusion, i.e. averaging scores of classifiers applied independently on the respective featuresets. This proved to produce more reliable results than early fusion.

4 EXPERIMENTAL RESULTS

All our evaluations are based on per-interaction accuracy of emergent leadership predictions as in [22, 23]. Specifically, an interaction is counted as correct, if and only if predicted and ground truth emergent leader coincide.

Offline Prediction

To evaluate the extent to which classifiers trained on a source dataset are able to achieve high performance on a target dataset, we train on PAVIS and test on MPIIGroupInteraction. At test time we assume to have access to a full test recording, i.e. we are predicting emergent leadership after an interaction took place ("offline" setting). In order to ensure using the same length for each of the approximately 20 minute long interactions on MPIIGroupInteraction we always use the first 19 minutes for feature extraction.

Figure 2 shows the obtained results for different feature sets and source- and target dataset combinations. The highest performance in the cross-dataset setting ("Source: PAVIS, Target: MPI") is achieved by a combination of VFOA and pose features with an accuracy of 0.68, slightly outperforming VFOA features only at 0.64 accuracy. Combining other featuresets (e.g. face) with VFOA and pose did not improve results, therefore we do not show these combinations in Figure 2. In case video recordings are not available or desired, an accuracy of 0.5 can be achieved with speaking activity features only. Both results are clearly above the random baseline of 0.29, showing the feasibility of cross-dataset prediction.

Comparing cross-dataset to within-dataset results reveals that cross-dataset accuracies are consistently lower than within-dataset accuracies on PAVIS. More surprisingly, by

Feature	MPI		PAVIS	
	Acc.	Ori.	Acc.	Ori.
totWatcherNoME	0.59	+	0.66	+
ratioWatcherLookSOne	0.59	+	0.62	+
totWatcher	0.55	+	0.76	+
maxTwoWatcherNoME	0.45	+	0.21	+
minTwoWatcherWME	0.45	−	0.14	+
minTwoWatcherNoME	0.41	−	0.14	−
totME	0.36	+	0.60	+

Table 1: Accuracies for single feature based classification using selected VFOA features on PAVIS and MPIIGroupInteraction. "Ori." indicates whether the maximum or the minimum of the feature was used for prediction.

training on PAVIS, we achieve higher accuracies on MPIIGroupInteraction compared to training on MPIIGroupInteraction directly. This is most likely an effect of the limited training data available on MPIIGroupInteraction. In total there are only 78 samples (one per participant), compared to 232 samples on PAVIS due to the segment based annotations.

Within datasets, we achieve the best accuracy for PAVIS with a combination of speaking activity, VFOA and pose features (0.86). The best result for the emergent leadership detection task on PAVIS, published in [7], achieved detection scores of 0.76 for the positive and 0.93 for the negative class with a combination of pose and VFOA features. Later work by the same authors adopted a different evaluation setting, and thus can not serve as a comparison [8, 9]. The detection scores for our predictions on PAVIS based on VFOA, pose and speaking activity features reach 0.86 for the positive and 0.95 for the negative class, exceeding the previously published results. Likely as a result of fewer training examples, within-dataset results on MPIIGroupInteraction are much lower, with a maximum accuracy of 0.45 for VFOA features.

Online Prediction

Some applications scenarios require information about emergent leaders already during the course of an interaction. To evaluate in this setting, we restrict the time interval from which to extract features from the test interactions. Figure 3 shows accuracies for classifiers that only observe data from a limited number of minutes at the beginning of the interaction. Both our best performing featureset (VFOA and pose) and speaking activity features tend to achieve higher accuracies after longer observation time. This tendency is more pronounced for the VFOA and pose featureset, which stays between 0.4 and 0.6 accuracy during the first minutes of an interaction, and clearly above 0.6 accuracy after more than 15 minutes. Thus, while prediction above chance is possible early on, longer observation is required for optimal precision.

Feature Analysis

VFOA features were the best performing individual featureset in our evaluation. To better understand which VFOA features generalise best across datasets, we quantify how well each individual feature discriminates the ground truth classes on MPIIGroupInteraction and PAVIS. For each feature, we define an unlearned classifier that simply selects the person with either the maximum or the minimum value on that feature as the emergent leader of an interaction. We decide on selection via minimum or maximum based on which strategy achieves higher accuracy. We refer to features of which we take the maximum/minimum as having positive/negative orientation respectively. This is not a valid classification approach, as we do not employ cross-validation. Instead, it is a post-hoc analysis on the connection between individual features and ground truth. See Table 1 for the features with accuracy of at least 0.5 on both datasets (informative and good transfer) along with the features showing a difference of at least 0.2 accuracy between both datasets (weak transfer). Find the full table in the supplementary material. The features with the highest accuracies on both datasets are *totWatcher* (total time a person is watched by others), *totWatcherNoME* (totWatcher given there is no mutual eye contact (ME)) and *ratioWatcherLookSOne* (ratio between totWatcher and the time a person looks at other people). This indicates that being looked at by others is a central property of leaders on both datasets. In contrast, the low performance of *totME* on MPIIGroupInteraction in comparison to the high performance on PAVIS indicates that mutual eye contact is less robustly associated with leadership across the two datasets. The accuracy of *maxTwoWatcherNoME*, *minTwoWatcherWME* and *minTwoWatcherNoME* (the max/min time a person is looked at by two others while having/not having ME) differs strongly between the datasets while always staying below 0.5.

5 CONCLUSION

In this paper, we were first to investigate a cross-dataset evaluation setting for the emergent leadership detection task. We showed that it is possible to predict emergent leadership from nonverbal features on a new dataset not observed at test time, with a combination of VFOA and pose features achieving best performance. Furthermore, we analysed the feasibility of online prediction and the usefulness of single VFOA features. All in all, our initial study on cross-dataset emergent leadership prediction opens the way to investigate this important task in more realistic settings.

ACKNOWLEDGMENTS

This work was funded, in part, by a JST CREST research grant under Grant No.: JPMJCR14E1, Japan. We thank Cigdem Beyan for sharing her code for pose feature computation.

REFERENCES

[1] John E Baird Jr. 1977. Some nonverbal elements of leadership emergence. *Southern Speech Communication Journal* 42, 4 (1977), 352–361. https://doi.org/10.1080/10417947709372361

[2] Tadas Baltrušaitis, Marwa Mahmoud, and Peter Robinson. 2015. Cross-dataset learning and person-specific normalisation for automatic action unit detection. In *Proc. of the IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*, Vol. 6. 1–6. https://doi.org/10.1109/FG.2015.7284869

[3] Tadas Baltrusaitis, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency. 2018. Openface 2.0: Facial behavior analysis toolkit. In *Proc. of the IEEE International Conference on Automatic Face & Gesture Recognition*. 59–66. https://doi.org/10.1109/FG.2018.00019

[4] Cigdem Beyan, Francesca Capozzi, Cristina Becchio, and Vittorio Murino. 2016. Identification of emergent leaders in a meeting scenario using multiple kernel learning. In *Proc. of the Workshop on Advancements in Social Signal Processing for Multimodal Interaction*. 3–10. https://doi.org/10.1145/3005467.3005469

[5] Cigdem Beyan, Francesca Capozzi, Cristina Becchio, and Vittorio Murino. 2018. Prediction of the Leadership Style of an Emergent Leader Using Audio and Visual Nonverbal Features. *IEEE Transactions on Multimedia* 20, 2 (2018), 441–456. https://doi.org/10.1109/TMM.2017.2740062

[6] Cigdem Beyan, Nicolò Carissimi, Francesca Capozzi, Sebastiano Vascon, Matteo Bustreo, Antonio Pierro, Cristina Becchio, and Vittorio Murino. 2016. Detecting emergent leader in a meeting environment using nonverbal visual features only. In *Proc. of the ACM International Conference on Multimodal Interaction*. 317–324. https://doi.org/10.1145/2993148.2993175

[7] Cigdem Beyan, Vasiliki-Maria Katsageorgiou, and Vittorio Murino. 2017. Moving as a Leader: Detecting Emergent Leadership in Small Groups using Body Pose. In *Proc. of the ACM Multimedia Conference*. 1425–1433. https://doi.org/10.1145/3123266.3123404

[8] Cigdem Beyan, Vasiliki-Maria Katsageorgiou, and Vittorio Murino. 2019. A Sequential Data Analysis Approach to Detect Emergent Leaders in Small Groups. *IEEE Transactions on Multimedia* (2019). https://doi.org/10.1109/TMM.2019.2895505

[9] Cigdem Beyan, Muhammad Shahid, and Vittorio Murino. 2018. Investigation of Small Group Social Interactions Using Deep Visual Activity-Based Nonverbal Features. In *Proc. of the ACM Multimedia Conference*. 311–319. https://doi.org/10.1145/3240508.3240685

[10] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2018. OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. In *arXiv preprint arXiv:1812.08008*.

[11] Vanessa Urch Druskat and Anthony T Pescosolido. 2006. The impact of emergent leader's emotionally competent behavior on team trust, communication, engagement, and effectiveness. *Research on Emotions in Organizations* 2 (2006), 25–55.

[12] Sebastian Feese, Amir Muaremi, Bert Arnrich, Gerhard Troster, Bertolt Meyer, and Klaus Jonas. 2011. Discriminating Individually Considerate and Authoritarian Leaders by Speech Activity Cues. In *Proc. of the IEEE International Conference on Privacy, Security, Risk and Trust and IEEE International Conference on Social Computing*. 1460–1465. https://doi.org/10.1109/PASSAT/SocialCom.2011.209

[13] Fabiola H Gerpott, Nale Lehmann-Willenbrock, Jeroen D Silvis, and Mark Van Vugt. 2018. In the eye of the beholder? An eye-tracking experiment on emergent leadership in team interactions. *The Leadership Quarterly* 29, 4 (2018), 523–532. https://doi.org/10.1016/j.leaqua.2017.11.003

[14] Leonard D Goodstein and Richard I Lanyon. 1999. Applications of Personality Assessment to the Workplace: A Review. *Journal of Business and Psychology* 13, 3 (1999), 291–322. https://doi.org/10.1023/A:1022941331649

[15] Akko Kalma. 1992. Gazing in triads: A powerful signal in floor apportionment. *British Journal of Social Psychology* 31, 1 (1992), 21–39. https://doi.org/10.1111/j.2044-8309.1992.tb00953.x

[16] Jill Kickul and George Neuman. 2000. Emergent Leadership Behaviors: The Function of Personality and Cognitive Ability in Determining Teamwork Performance and KSAs. *Journal of Business and Psychology* 15, 1 (2000), 27–51. https://doi.org/10.1023/A:1007714801558

[17] Ahmet Alp Kindiroglu, Lale Akarun, and Oya Aran. 2017. Multi-domain and multi-task prediction of extraversion and leadership from meeting videos. *EURASIP Journal on Image and Video Processing* 2017, 1 (2017), 77. https://doi.org/10.1186/s13640-017-0224-z

[18] Philipp Müller, Michael Xuelin Huang, and Andreas Bulling. 2018. Detecting Low Rapport During Natural Interactions in Small Groups from Non-Verbal Behavior. In *Proc. of the ACM International Conference on Intelligent User Interfaces*. https://doi.org/10.1145/3172944.3172969

[19] Philipp Müller, Michael Xuelin Huang, Xucong Zhang, and Andreas Bulling. 2018. Robust Eye Contact Detection in Natural Multi-Person Interactions Using Gaze and Speaking Behaviour. In *Proc. of the International Symposium on Eye Tracking Research and Applications*. 31:1–31:10. https://doi.org/10.1145/3204493.3204549

[20] Shogo Okada, Laurent Son Nguyen, Oya Aran, and Daniel Gatica-Perez. 2019. Modeling Dyadic and Group Impressions with Intermodal and Interperson Features. *ACM Transactions on Multimedia Computing, Communications, and Applications* 15, 1s (2019), 13. https://doi.org/10.1145/3265754

[21] John Platt. 1999. Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. *Advances in Large Margin Classifiers* 10, 3 (1999), 61–74.

[22] Dairazalia Sanchez-Cortes, Oya Aran, Dinesh Babu Jayagopi, Marianne Schmid Mast, and Daniel Gatica-Perez. 2013. Emergent leaders through looking and speaking: from audio-visual data to multimodal recognition. *Journal on Multimodal User Interfaces* 7, 1-2 (2013), 39–53. https://doi.org/10.1007/s12193-012-0101-0

[23] Dairazalia Sanchez-Cortes, Oya Aran, Marianne Schmid Mast, and Daniel Gatica-Perez. 2012. A Nonverbal Behavior Approach to Identify Emergent Leaders in Small Groups. *IEEE Transactions on Multimedia* 14, 3 (2012), 816–832. https://doi.org/10.1109/TMM.2011.2181941

[24] R Timothy Stein and Tamar Heller. 1979. An empirical analysis of the correlations between leadership status and participation rates reported in the literature. *Journal of Personality and Social Psychology* 37, 11 (1979), 1993–2002. https://doi.org/10.1037/0022-3514.37.11.1993