Modeling Dominance Effects on Nonverbal Behaviors Using Granger Causality

Kyriaki Kalimeri
University of Trento, CIMEC
Bruno Kessler Foundation
kalimeri@fbk.eu

Bruno Lepri
Bruno Kessler Foundation
MIT Media Lab
lepri@fbk.eu

Oya Aran
Idiap Research Institute
oaran@idiap.ch

Dinesh Babu Jayagopi
Idiap Research Institute
djaya@idiap.ch

Daniel Gatica-Perez
Idiap Research Institute
Ecole Polytechnique Federal de Lausanne (EPFL)
gatica@idiap.ch

Fabio Pianesi
Bruno Kessler Foundation
pianesi@fbk.eu

ABSTRACT

In this paper we modeled the effects that dominant people might induce on the nonverbal behavior (speech energy and body motion) of the other meeting participants using Granger causality technique. Our initial hypothesis that more dominant people have generalized higher influence was not validated when using the DOME-AMI corpus as data source. However, from the correlational analysis some interesting patterns emerged: contradicting our initial hypothesis dominant individuals are not accounting for the majority of the causal flow in a social interaction. Moreover, they seem to have more intense causal effects as their causal density was significantly higher. Finally dominant individuals tend to respond to the causal effects more often with complementarity than with mimicry.

Categories and Subject Descriptors

I.5.1 [**Computing Methodologies**]: PATTERN RECOGNITION—*Models*

General Terms

Algorithms

Keywords

Social and Group Interaction, Multimodal Fusion and Integration, Granger Causality

1. INTRODUCTION

One of the basic mechanisms of social interaction and one of the fundamental dimensions for analyzing the group dynamics and the formation of a group social structure is dominance [6]. In social psychology, dominance is usually seen in two ways: (i) as a personality characteristic (a trait) [13] or (ii) a sign of a person's hierarchical position within a group [13]. Several social psychology studies have shown that individuals higher in trait dominance tend to attain more influence in face-to-face interactions [3, 17]. In the last years dominance aroused much interest in the domain of social interaction from sensor data [1]. Different researchers have dealt with the automatic detection of the most dominant person and/or of the least dominant person in small group interactions (e.g. meetings) using nonverbal acoustic and visual cues [10, 11]. Some studies in social psychology [19] confirm that people can respond to dominant behaviors with either mimicry or complementarity behaviors, where the former amounts to a reproduction of the behavior of the dominant person and the latter to an opposite behavior. Recently, some works have started to deal with the automatic detection of mimicry. In [18] the way in which visual and vocal behaviors displayed by two interlocutors can be used to detect and identify visual and vocal mimicry was investigated. On the visual side, they detected the existence of a correlation between the motion intensities of two interacting persons. They were also able to detect similar correlations between vocal features showing that people change their vocal style while interacting with others and that the change is in the direction of mimicry. However, according to Chartrand and Bargh [5], mere correlational approaches are not enough to conclude that person X is mimicking (or complementing) person Y; rather, they can only inform whether X and Y are displaying similar or contrasting behavioural patterns. In order to conclude for the presence of true mimicry/complementarity, a causal relationship must be proven in which the display of a particular behaviour and then person Y mimics (or complements) that behaviour.

In this work, our goal is to automatically model the causal effects that people displaying dominant nonverbal behaviors have on the nonverbal behaviors of the other participants. In order to investigate these effects, we apply Granger causality, an approach that detects and estimates the direction of causal influence in time series analysis. To exemplify this approach, in this work we focus on peoples' nonverbal activity, both vocal and kinesic, detected by means of acoustic

ICMI'12, October 22–26, 2012, Santa Monica, California, USA.

and visual cues and the way that affects the nonverbal behavior of other group members. Granger causality [9] is a promising approach to this end: widely used in neuroscience to infer the existence of causal relationships among neural circuits and model causal relationships among temporal series. To our knowledge, it has been seldom applied to the automatic analysis of human behavior [14, 12] and to social behavior, in particular.

2. THE DOME CORPUS

In our work we exploited a multimodal corpus, DOME, which includes small group conversations with dominance annotations [2]. It contains five-minute non-overlapping slices selected from a subset of meetings of the popular Augmented Multi-party Interaction (AMI) corpus. The DOME corpus has been presented in [2] and used in previous studies for the modeling of dominance based on automatically extracted audio-visual nonverbal cues [1, 11]. Each meeting has four participants. Meetings in the AMI corpus were carried out in a multi-sensor meeting room which contained a table for four participants, a slide screen and a white board. The audio was recorded via several microphones: two circular microphone arrays on the ceiling and on the table, headset and lapel microphones. The video was recorded via seven cameras (see Figure 1 for sample screen shots).

Figure 1: Screen shots from the corpus from the available close-up cameras.

The entire DOME corpus consists of two subsets of meetings, corresponding to 10 hours of meeting data. Following the "thin slice" approach, every meeting has been divided in five-minute non-overlapping meeting segments. The first subset, M1, which is used in this study, contains 58 five-minute meeting segments. The segments were selected from 11 scenario meetings in AMI corpus.

Each meeting segment is annotated by three annotators, who used their own judgements on dominance to rank the meeting participants. No prior definition of dominance is given to the annotators. Each annotator ranked each meeting participant from 1 to 4, with 1 representing the most dominant person, and 4 representing the least dominant person in the meeting. The annotations have been analyzed to assess the agreement between the annotators for the most and least dominant participants. A detailed analysis of the corpus and the annotations can be found in [2].

3. AUDIO-VISUAL CUES EXTRACTION

A solid body of work in social psychology and social computing has documented the role that nonverbal communicative cues play in the expression and perception of dominant behavior [13, 6]. We focus on peoples' nonverbal activity detected by means of acoustic and visual cues and the way that affect the nonverbal behavior of other group members. To this end, we extracted two features: speaking energy and body motion activity.

3.1 Acoustic Cues

The speaking energy was extracted from the four close-talk microphones attached to each of the meeting participants, one per person. In particular, we computed a speaker energy value for each participant using a sliding window at each time step as described in [21]. The value of speaking energy was extracted using the root mean square amplitude of the audio signal over a sliding time window for each audio track. A window of 40 ms was used with a 10 ms time shift.

3.2 Visual Cues

To estimate the nonverbal kinesic behavior of the participants, we have used compressed domain processing [20] in order to extract the motion from the skin colored regions. The motion vectors and residual coding bit rate features are extracted from compressed domain videos. A motion vector of a source block in frame t indicates which predictor block from frame $t-1$ is to be used. The extracted motion vectors are further filtered and for each motion vector a confidence measure is computed by using DCT coefficients that measure the amount of local texture and only the vectors with high confidence are kept. In order to capture finer motion, such as moving lips, etc., we use the residual coding bitrate. After motion compensation, the DCT coefficients of the residual signal, which is the difference between the block to be encoded and its prediction from the reference frame, are quantized and entropy coded. The residual coding bitrate is the number of bits used to encode this transformed residual signal.In combination with the motion vector magnitude, the residual coding bitrate provides complementary evidence for visual activity. The skin-colored blocks in the compressed domain are detected using a Gaussian mixture model to identify hand and head regions of participants, and motion features are computed only on these blocks. We further normalized the motion vector magnitudes and residual coding bitrate, with respect to the average participant and overall activity in the meeting [11]. The final feature representing the amount of motion is computed as the average of normalized motion vector magnitude and residual coding bit rate, and indicates the compressed domain activity levels for the participant for each time frame, with 25 fps frame rate.

4. MODELING DOMINANCE EFFECTS

To understand the direction of the influence flow in social interactions, it is of fundamental importance to distinguish the driver from the recipient. One of the most prominent methods to estimate the direction of the causal influence in time series analysis is the Granger Causality(GC) [9]. This method is based on asymmetric prediction accuracies of one time series on the future of another. In detail, let two time series X_1 and X_2 be defined as,

$$X_1(t) = \Sigma_{j=1}^{p} A_{11,j} X_1(t-j) + \Sigma_{j=1}^{p} A_{21,j} X_2(t-j) + \xi_1(t) \quad (1)$$

$$X_2(t) = \Sigma_{j=1}^{p} A_{21,j} X_1(t-j) + \Sigma_{j=1}^{p} A_{22,j} X_2(t-j) + \xi_2(t) \quad (2)$$

where A is the matrix containing the coefficients of the model and ξ_1,ξ_2 are the residuals of X_1 and X_2 respectively. A time series X_1, is said to Granger-cause X_2 if the inclusion of past observations of X_1 reduces the prediction error of X_2 in a linear regression model of X_2 and X_1, as compared to a model including only the previous observations of X_2. An important aspect of GC is its generalizability to the multivariate case in which the GC of X_1 on X_2 is tested in the context of multiple additional variables (in our scenario the other two meeting participants denoted by W and Z). In this case, X_1 is said to Granger-cause X_2 if knowing X_1 reduces the variance in X_2's prediction error when all the other variables are also included in the model [8].

In our case we defined two systems, one in which the time seriesX_1, X_2, X_3, X_4 of the system $\mathbf{X}$ refer to the body movement of each of our subjects and a second one in which the time series refer to their speaking activity as described above. To remove every linear trend from the data, all series have been de-trended and their temporal mean has been removed as an initial preprocessing step. We estimated the best order of the multivariate autoregressive model (MVAR) using the Bayesian Information Criterion (BIC) [15]. The estimated model was further checked both (i) to control whether it accounted for a sufficient amount of variance in the data and (ii) using the Durbin-Watson [7] test to validate whether its residuals are serially uncorrelated. Then, once the set of significant lagged values for X_2 is found, the regression is augmented with lagged levels of X_1. Having estimated the G-causality magnitudes, their statistical significance was evaluated via an F-test on the null hypothesis that the coefficients $A_{i,j}$ are zero. If the coefficients in the corresponding $A_{i,j}$ were jointly significantly different from zero, then the causal interaction was considered to be statistically significant. To correct the tests from multiple comparisons, the Bonferroni correction [4] approach was chosen thresholded at $\dfrac{P}{n(n-1)}$, with P=0.01.

Let our group of participants be a small causal network of four interacting nodes. In causal networks, the nodes represent variables and the directed edges represent causal interactions. A measure of the causal interactivity of a system $\boldsymbol{X}$ is the causal density [16], which is defined as the mean of all pairwise G-causalities between system elements, conditioned on the system's statistically significant interactions.

$$cd(X) \equiv \frac{1}{n(n-1)} \Sigma_{i \neq j} F_{X_i \rightarrow X_j | X_{[ij]}}$$

where $X_{[ij]}$ is the network from which the variables X_i and X_j are omitted. For each of our nodes (i.e. each subject), we estimate the unit causal density $cd_u(i)$ which is the summed causal interactions involving a node i normalized by the number of nodes. Furthermore, to identify nodes with distinctive causal effects on the network dynamics, we estimated the *causal flow* of a subject X_1. The causal flow is defined as the difference between the in-degree and the out-degree of a given node. Therefore, a subject with a high positive causal flow exerts a strong causal influence on the meeting and it can be called a causal *source.* On the other side, a subject with a highly negative causal flow can be called a causal *sink.* From the GC relationships in the causal network, we are only able to determine if the speech/body activity of subject X_1 has a causal effect on the speech/body activity of subject X_2; however, we are not able to discriminate between mimicry and complementarity effects.

5. RESULTS

Initially, we focus our attention on the relationships between influence, behavior and the dominance scores. Our expectation was that more dominant people have generalized higher influence, measured in terms of higher density, positive and higher flow and higher out-flow. For both modalities we computed the Spearman rank correlation between the causal flow and the dominance scores obtained by the 3 annotators. The four participants were ranked according to their score for the different measurements on a slice by slice time basis. In Figure 2, both the distributions for the bodily activity (blue bars), the speech activity (red bars) and the combination of the normalized causal effect of the two modalities (green bars) suggest a tendency of the most dominant people to adapt to the nonverbal behavioral manifestation of the other participants. This evidence is against our initial hypothesis as we expected the less dominant participants to manifest this kind of behavior.

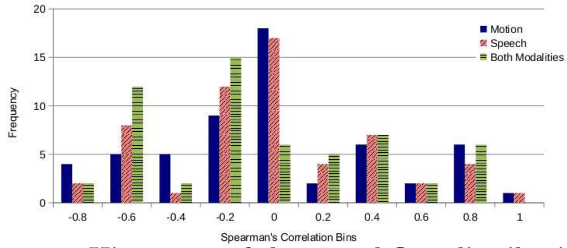

Figure 2: Histogram of the causal flow distribution. From left to right: In blue the bodily activity, in red the speaking activity and in green the joint contribution correlations

Similarly, for both modalities we computed the Spearman rank correlation between the causal density and the dominance scores obtained by the 3 annotators. Again, all ranks are computed on a slice by time slice basis. In Figure 3, the pattern of the causal density has interestingly changed, showing an increase of the positive correlations for both modalities.

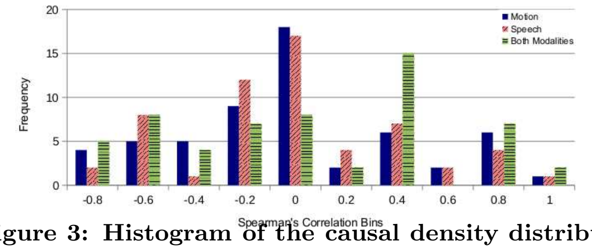

Figure 3: Histogram of the causal density distribution. From left to right: In blue the bodily activity, in red the speaking activity and in green the joint contribution correlations

In both Figure 2 and Figure 3 the peak of the distributions is zero because of the fact that we have considered only the measurements that were statistically significant considering all the others to be zero. As emerges from the correlations between the causality flow and the dominance annotation ranking the most dominant people are not usually the ones that account for the highest quantity of causal flow, in contrast with our initial hypothesis. In order to assess the phenomena of mimicry and complementarity, we investigated the correlation between the time series of the subjects for which we found some significant causal effect. For example, once determined that subject X_1 Granger-causes X_2, we checked if the histogram of the correlation between the time series X_1 and the time series X_2 is positive, possibly suggesting mimicry effects, or is negative, suggesting complementarity ones. The patterns that emerged showed a tendency of the most dominant individuals to reply with complementarity as in most of the cases the correlation was negative. Out of 51 cases we classified 20 mimicry responses and 31 complementarity ones.

6. CONCLUSIONS

In this paper we have employed Granger Causality in the analysis of the effects dominant behavior might on the nonverbal behavior of others in a small group interaction and we further characterized them according to their potential relation to mimicry and complementarity patterns. Using the DOME meeting corpus, the patterns emerged from the correlational analysis are interesting, contradicting our initial hypothesis that dominant individuals do not appear to account for the majority of the causal flow in social interactions, confirming the findings of [12] obtained from the analysis of accelerometer data on a dataset of brainstorming and problem solving tasks. Moreover, dominant individuals seem to have more intense causal effects as their causal density was significantly higher while they tend to respond to the causal effects more often with behavior we hypothesize to be related to complementarity rather than mimicry. The importance of this study lies on the novelty of the proposed model that takes into consideration the causal effects dominant subjects' nonverbal behavior has on the behavior of the other parties of a social interaction, rendering it a promising approach to the potential assessment of more complex behaviors. However, further analysis needs to be carried out in order to confirm the validity of these findings.

7. ACKNOWLEDGEMENTS

Bruno Lepri's research was funded by the Marie Curie-COFUND-7th Framework PERSI project. Oya Aran's research was funded by the Swiss National Science Foundation Ambizione project SOBE ($PZ00P2\_136811$). D. Jayagopi and D. Gatica-Perez thank the support of the SNSF NCCR IM2.

8. REFERENCES

[1] O. Aran and D. Gatica-Perez. Fusing Audio-Visual Nonverbal Cues to Detect Dominant People in Group Conversations. *2010 20th International Conference on Pattern Recognition*, 0:3687–3690, 2010.

[2] O. Aran, H. Hung, and Gatica-Perez. A Multimodal Corpus for Studying Dominance in Small Group Conversations. In *LREC workshop on Multimodal Corpora, Malta (LREC MMC'10)*, 2010.

[3] E. J. Aries, C. Gold, and R. H. Weigel. Dispositional and situational influences on dominance behavior in small groups. *Journal of Personality and Social Psychology*, 44(4):779–786, 1983.

[4] Y. Benjamini and Y. Hochberg. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300, 1995.

[5] T. L. Chartrand and J. A. Bargh. The chameleon effect: the perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, 76(6):893–910, 1999.

[6] N. E. Dunbar and J. K. Burgoon. Measuring nonverbal dominance. *The sourcebook of nonverbal measures going beyond words*, page 361, 2005.

[7] J. Durbin and G. Watson. Testing for serial correlation in least squares regression.I. *Biometrika*, 37(3):409–428, 1950.

[8] J. Geweke. Measurement of Linear Dependence and Feedback Between Multiple Time Series. *Journal of the American Statistical Association*, 77(378):304–313, 1982.

[9] C. W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438, 1969.

[10] H. Hung, D. Jayagopi, C. Yeo, G. Friedland, S. Ba, J.-M. Odobez, K. Ramchandran, N. Mirghafori, and D. Gatica-Perez. Using audio and video features to classify the most dominant person in a group meeting. In *Proceedings of the 15th international conference on Multimedia*, pages 835–838. ACM Press, 2007.

[11] D. B. Jayagopi, H. Hung, C. Yeo, and D. Gatica-Perez. Modeling Dominance in Group Conversations Using Nonverbal Activity Cues. *IEEE Transactions on Audio, Speech {&} Language Processing*, 17(3):501–513, 2009.

[12] K. Kalimeri, B. Lepri, T. Kim, F. Pianesi, and A. Pentland. Automatic modeling of dominance effects using granger causality. In *HBU*, pages 124–133, 2011.

[13] M. S. Mast. Dominance as Expressed and Inferred Through Speaking Time. *Human Communication Research*, 28(3):420–450, 2002.

[14] K. Prabhakar, S. M. Oh, P. Wang, G. D. Abowd, and J. M. Rehg. Temporal causality for the analysis of visual events. In *CVPR*, pages 1967–1974, 2010.

[15] G. Schwarz. Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2):461–464, 1978.

[16] A. K. Seth. A MATLAB toolbox for Granger causal connectivity analysis. *Journal of Neuroscience Methods*, 186(2):262–273, 2010.

[17] J. A. Smith and R. J. Foti. A pattern approach to the study of leader emergence. *The Leadership Quarterly*, 9(2):147–160, 1998.

[18] X. Sun, J. Lichtenauer, M. F. Valstar, A. Nijholt, and M. Pantic. A multimodal database for mimicry analysis. In *Proceedings of the 4th Bi-Annual International Conference of the HUMAINE Association on Affective Computing and Intelligent Interaction (ACII2011)*, Memphis, Tennessee, USA, October 2011.

[19] L. Z. Tiedens and A. R. Fragale. Power moves: complementarity in dominant and submissive nonverbal behavior. Technical Report 3, Graduate School of Business, Stanford University, California 94305-5015, USA. ltiedens@stanford.edu, 2003.

[20] C. Yeo, P. Ahammad, K. Ramchandran, and S. Sastry. High-speed action recognition and localization in compressed domain videos. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(8):1006–1015, 2008.

[21] D. Zhang, D. Gatica-Perez, S. Bengio, and I. McCowan. Modeling individual and group actions in meetings with layered HMMs. *IEEE Transactions on Multimedia*, 8(3):509 – 520, 2006.