To appear: Special Session on Interactive Behavior Analysis in
5th International Symposium on Communications, Control, and Signal Processing

Modeling Conversational Dynamics and Performance in a Social Dilemma Task

Wen Dong[1], Bruno Lepri[1,2], Taemie Kim[1], Fabio Pianesi[2], and Alex "Sandy" Pentland[1]
[1]MIT Media Lab
Cambridge, MA 02124
[2]FBK
Povo-Trento, Italy
Email: {wdong,kim,pentland}@media.mit.edu, {lepri,pianesi}@fbk.eu

Abstract— **In this paper, we describe a novel approach, based on Markov jump processes, to model group interaction dynamics and group performance. In particular, we estimate conversational events such as turn taking, backchannels, turn-transitions at the micro-level (1 minute windows) and then we bridge the micro-level behavior and the macro-level performance. We test this approach with a cooperative game dataset and we verified the relevance of micro-level interaction dynamics in determining a good group performance (e.g. higher speaking turns rate and backchannels rate and lower turns competition rate).**

Keywords – group performance; small group dynamics; nonverbal behavior; stochastic modeling.

I. INTRODUCTION

Nowadays, management, scientific research, politics and a lot of other activities are accomplished by groups. For this reason, it is becoming even more important to understand the determinants of group performance. The research area of organizational behavior has proposed and tested methods to improve the effectiveness of group collaboration and to deal with the problem of group suboptimality, groups tend to perform better than individuals but not as well as they could [6]. In particular, group dynamics have been one of the focuses as it is a key factor affecting the performance and the satisfaction of the group [12].

For instance, Hall and Watson [2] demonstrated that the performance of a group is noticeably affected by the understandings from its members on what is a productive group process, and that the group performance could be improved by just instructing the group members to be more participative and engaged in the conversation. According to them [2], a more productive group is more likely to generate group answers that are better than the individual answers by reconciling the differences among its members with win-win strategies and through 'aha' experiences. Wilson et al. [13] observed several tens of group processes in solving two versions of the 20-questions game. They noted that (1) groups solve significantly larger proportions of the games than individuals, (2) the questions asked by groups work increasingly better than those asked by individuals as a game proceeds and becomes harder, (3) a pair of strangers generate more (unique) ideas — that are compatible with a given list of yes/no questions and their answers — than a pair of friends, and a pair of friends generate more ideas than two individuals working alone. Many issues related to the lacking of participation, such as social-loafing and production-blocking, have been discussed by various researchers [4]. A very recent and interesting study has shown that groups perform better on tasks if the members have strong social skills and if the conversation reflects more group members' ideas [14]. The tasks could range widely from brainstorming to quantitative analysis to negotiation and are drawn from all the quadrants of the McGrath Task Circumplex [8], a well-established taxonomy of group tasks based on the coordination processes they require. The major findings were that group performance is not related to the average or maximum of the members' performances but it is correlated with average social sensitivity of group and with the equality in distribution of turn taking. This and many previous findings support the speculation that certain aspects of the interactions among the members are important to group performance and are independent of specific tasks.

In our paper we propose and discuss a novel approach to relate the microcosmic interaction patterns among the group members to the group performance. More precisely, we propose to use the Markov jump process model, an extension of Markov chains when time is considered to be continuous instead of discrete, in order to capture how the microcosmic interaction patterns will generate the macrocosmic interaction statistics such as equal participation among the group members and engagement, which in turn will have consequences on group performances.

In particular, we focus on modeling the form of the interaction and conversation in small group meetings. Our proposed Markov jump process models the turn taking in small group conversations following the turn taking *systematics* proposed by Sacks et al. [11]. Roughly speaking, the turn taking *systematics* consists of turn-constructional features for determining where transition will be relevant, two types of turn-allocational techniques (current speaker selects the next one and self-selection) for determining how a next turn will be allocated, and a set of practices for employing the turn-allocational techniques by reference to transition-relevance places. In sum, in this model the current speaker selects the next speaker, the next speaker self-selects himself or the current speaker continues at transition places. Sacks et al. used this simple systematic to explain how the conversations are locally managed, party administrated, interactively controlled and sensitive to recipient design.

We believe tracking group performance through tracking the turn taking behavior from signals such as audio variance, motion, and who-faces-whom has the following advantages. First, it is computationally cheap and power efficient. Second, it combines multiple types of sensor data in a unified framework and achieves better performance by related different types of signals. Third, it tells us how microcosmic interaction can have macrocosmic consequence on performance.

II. OUR MODEL

We use Markov jump process, an extension of Markov chains when time is considered to be continuous instead of discrete, to estimate conversational turns by using the following multimodal cues: (i) speech variance, (ii) body movement variance collected using a 3 axes accelerometer, (iii) who faces whom by means of infrared scanning The rationale of using not only speech variance but also body movement variance and information about face-to-face interactions is based on some background literature. In Kendon [5] was showed that the addresser-addressee pair can be easily determined by who faces whom. Then, Harrigan [3] found that the amount of listeners' bodily activation is correlated with the speaking activity of the speaker and have a relevant impact on the conversational dynamics (e.g. turn-taking).

Markov jump process is likely to output that a person is speaking if his recorded audio intensity is greater than an estimated threshold, and we carefully adjust the thresholds of the persons in a group with an optimization algorithm so that the turn-taking structure is maximally satisfied. The audio intensity for an individual in a group discussion is assumed to be a linear combination of the audio intensities of all individuals in the discussion, and the intended individual has more contribution to the intensity. In our framework, we define a *speaking turn* as one continuous segment, not less than 1.5 sec., where a participant starts and ends her/his speech. Then, we modeled the following aspects of the turn-taking structure: (i) *taking the turn*: if nobody is taking the turn, then somebody should take the turn; (ii) *backchannel* [15]: we define backchannel as the situation where a subject Y speaks after a subject X for less than 1 sec. (e.g. "yes" or "uh-huh"); (iii) *speaker transitions*: if somebody is ending the turn, then she/he will transfer to another person. Roughly speaking, we have a *speaker transition* instead of a *taking the turn* when a the speaking turn of a subject Y follows in systematic way the speaking turn of a subject X ; (iv) *turn competition*: if two persons competing for turn, then one person will win. We define a *turn competition* as a situation in which 2 subjects are speaking at the same time and one ends before the other.

Specifically, the conversational state consists of whether speakers have turns. The conversational state of a group of C persons at time t is expressed as a **state vector** $x(t) = (person\ 1\ not\ speaking?\ person\ 1\ speaking? \vdots, \cdots, \vdots\ person\ C\ speaking?\ person\ C\ not\ speaking?)'$. In general, elements of $x(t)$ can contain any value besides Boolean values, such as the number of chemicals in chemistry, the number of species in ecology and the price of an asset in economy.

Conversational state $x(t)$ is changed by different events $(1, \cdots, v)$, and it also determines the rates $h_v\big(x(t)\big)$ at which different events will happen. We use **event vector** to describe the number of different events happening in a time window: $r = (r_1, \cdots r_v)'$ where r_i is the number of events of type i. We denote an event by a "reaction" $\sum_i \alpha_i x_i \rightarrow \sum_j \beta_j y_j$ where α_i number of reactant x_i has been consumed and β_j number of product has been generated. In our model of conversational dynamics, an event moves turn-taking status, and α_i, β_j are all one. We care about 4 types of events in our modeling: taking a turn, yielding a turn, transferring a turn and speaking in a back channel. We used Bayesian priors to bias the event rates towards reasonable values and tune the hyper-parameters of the priors manually. We considered 36 events in our analysis of four-person conversations: 4 different rates for the four persons to take a turn when nobody is currently taking the turn, 4 rates to yield a turn, 4×4 rates to transfer turn, 4 rates to speak in a back channel, 4 rates to seize a turn when another one is having the turn, and 4 rates to yield a turn when two or more persons try to take their turns simultaneously.

We use matrix algebra to express how events change conversational state. To this end we define the **reaction matrix** A as a $C \times v$ matrix where C is the length of the state vector $x(t)$ and v is the number of reactions. An element at column k and row j represents the amount added to state $x_j(t)$ if reaction k happens. In our modeling of conversational dynamics, entries of A are either $+1$ or -1 representing moving into a state or moving out of a state. For example, in the following equation, the first three columns of A represent speaker 1 starts to speak, speaker 2 transfers turn to speaker 3, and speaker 4 stops speaking. The column vector r means a speaker-transition event has happened. If we multiply A by r, we get an update the state matrix.

$$A \cdot r = \begin{pmatrix} -1 & & & \\ 1 & & & \\ & 1 & & \\ & -1 & & \cdots \\ & -1 & & \cdots \\ & 1 & & \\ & & 1 & \\ & & -1 & \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix} = \Delta x$$

Let $r(t_i)$ be the event vector representing the numbers of different events happening between t_i and t_{i+1}. In the ideal situation $r_{v_i}(t_i) = 1$ and the other elements of $r(t_i)$ are 0 because only event v_i happened during the period. The system states starting from $x(t_0)$ and corresponding to the sequence of events are updated according to $x(t_{i+1}) = x(t_i) + A \cdot r(t_i)$.

In order to derive the inference algorithm for estimating turn-taking dynamics from noisy sensor data, we begin with the ideal situation that we know all events (t_i, v_i), where $i = 1, \cdots, n$, $0 = t_0 < t_1 < \cdots < t_n =$T and $v_i \in \{1, \cdots, v\}$. The probability for this sequence of events to happen is

$$\mathrm{P}(v,x) = \prod_i h_{v_i}(x_t) \cdot \exp\left(-\sum_i h_{v_i}(x_t) \cdot (t_{i+1} - t_i)\right)$$

In reality we only have discrete time observations $y(n \cdot \Delta t)$ such as audio variance, body movement variance and detection of face-to-face configuration, and we want to infer from these discrete time observations how many, when and what events happened between these observations. The inference algorithm becomes non-trivial when the time interval between two consecutive observations becomes large, when we have missing data, and when we have data that are incompatible with the model. However it is possible to construct exact MCMC algorithms for inference based on discrete time observations [1], and it is possible to make inference with mean field approximation and variational method [9].

We introduced the following approximations to make the inference of turn-taking dynamics conceptually much simpler. Our first approximation is that turn-taking events only happen at the times of observation, and this approximation introduces 0.05 second error in the event times. Our second approximation is that at most one event can happen between two consecutive observations or 0.1 second. Our third approximation is that the observations for inferring turn-taking state have joint Gaussian distributions conditioned on turn-taking state.

Thus the probability of a sequence of latent events $v(t)$, together with the corresponding latent states $x(t)$ and observations $y(t)$ is

$$P(v,x,y) = \prod_{t,c} P\left(y^{(c)}(t) \middle| x^{(c)}(t)\right) \cdot P(x(0)) \prod_t P\left(v(x(t))\right),$$

$$P\left(v(x(t)) = 0\right) = \exp\left(-\sum_j h_j\,(x_t)\right),$$

$$P\left(v(x(t)) = i\right) = \frac{h_i(x(t))}{\sum_j h_j(x(t))}\left(1 - \exp\left(-\sum_j h_j(x(t))\right)\right),$$

$$P\left(y^{(c)}(t) \middle| x^{(c)}(t)\right) = N\left(\mu_{x^{(c)}(t)}, \Sigma_{x^{(c)}(t)}\right).$$

We use Gibbs sampling to infer latent states and parameters:

$$v(x(t))|\mu,\Sigma \sim Categorical\left(\left\{P\left(v(x(t)) = i\right) \cdot P\left(y(t)|x(t)\right) P\left(v(x(t+1))\right) : i = 0,\cdots,v\right\}\right),$$

$$h_i \sim Dirichlet\left(h + \sum_t \delta_{v(x(t)),i}\right),$$

$$\Sigma|\kappa_0,\mu_0,\nu_0,\Psi_0 \sim W^{-1}\left(T + \kappa_0, \Psi_0 + \sum_t (y_t - \bar{y})(y_t - \bar{y})^T + \frac{n\nu_0}{n+\nu_0}(\bar{y} - \mu_0)(\bar{y} - \mu_0)^T\right),$$

$$\mu|\Sigma,\nu_0,\mu_0 \sim N\left(\frac{n\bar{x} + \nu_0\mu_0}{T + \nu_0}, \frac{\Sigma}{n+\nu_0}\right).$$

III. Social Dilemma Task

We tested our Markov jump process framework on a dataset we collected to examine the relationship between the communication pattern and the group performance during a cooperation task, the Social Dilemma task. This dataset was collected from 50 groups of four members each, for a total of 200 participants. Each participant wore a Sociometric badge [9], a wearable electronic device with multiple sensors (e.g. microphone, infrared, accelerometer) able to detect face-to-face interactions, physical proximity, body movement data and speech features. Regarding the speech, due to privacy concerns the badges do not collect content of the speech or any other feature that may identify the speaker. The Figure 1 shows an example of participants wearing sociometric badges.

Figure 1: Example of an interacting group wearing sociometric badges around the neck.

The Social Dilemma task is a mixed-motive task often used in social psychology to measure the level of cooperation [7]. A social dilemma can be defined as a situation in which a group of people must decide between maximizing selfish interests and maximizing collective interests. It is generally more profitable for the individual to maximize selfish interests, but if all do so, all are worse off than if everyone had maximized collective interests. The performance of the group is measured as the sum of earnings of all members. So, this game does not depend on private or public information or even on facts per se, but rather on assessing the intentions of the other game players.

In our scenario, each subject was given sixteen $1 bills, of which they could invest any partial amount to the group fund and keep the rest for themselves. The total money gathered in the group fund was increased 50% by the experimenters and divided equally among the four participants regardless of how much each subject invested. Hence, investing all $16 to the group fund would help increase collective interests, but keeping all $16 would maximize personal earnings given a set of decisions of the other participants. However, if all participants were to keep all their money, all participants would earn 50% less than what they would have earned if everyone invested all. The instruction sheet had a sample calculation table and the experimenter confirmed that all subjects comprehended the situation and the consequences of their decisions.

IV. Experimental Results

The group performance of the social dilemma task is measured in terms of the contribution of the individuals to the group, which will be multiplied and shared with the whole group. Individuals need sufficient discussion to build mutual trust, because individuals need to make maximum contribution to the group to maximize the average individual gain, and selfish individuals will gain more assuming all other individuals behave similarly. The maximum total contribution to a group is \$64, and in reality we have total contributions ranging from \$16 to \$64 and averaging at \$51.15.

The event rates from our Markov model are useful in clarifying the relationship between the group discussion dynamics and the trust among members in the social dilemma experiment. On average, 2 additional turns per minute or 1 additional speaker transition per minute increase group money by \$1. In particular, all the groups with less than 25 speaking turns per minute have individuals holding back half of the money. We use the following log-linear model to capture the relationship between group money and the rate in 1 minute windows of four conversational events (turn taking, turn competitions, backchannel, turn taking by different subjects): $\log E(group\ money) = \alpha_0 + \sum_i \alpha_i \times rate\ of\ event\ i + Poisson\ error$. The results of our analysis are shown in Table 1.

Table 1: More active group discussion and more balanced discussion participation among the members indicate better performance in Social Dilemma task.

Performance percentile	Turn taking per minute	Turn competitions per minute	Back-channel per minute	Turn taking by different members
25%	25	2	5	5
50%	30	3	20	10
75%	47	3	25	17

In other words, individual turn-taking events accumulatively add to the group money, and the turn-taking events have "diminishing returns". Among the different events, turn taking rate and back channel rate increase group money, turn competition rate and uneven turn transferring decreases group money, and the four factors explain 37% variance of group money.

V. CONCLUSION

The aim of this paper was to investigate and to test a novel approach to reason about group performance through modeling and sensing small group interaction dynamics. We propose a Markov jump process framework to estimate conversational events such as turn taking, turn transitions, competition to take the turn and back channel looking at 1 minute windows, bridge micro-level non verbal behaviors and macro-level performance. We tested our model on a Social Dilemma task and we found that higher turn taking rate and backchannel rate increase the group performance (more money put for the group) while higher turn competition rate and uneven turn transferring decrease the group performance.

On the practical side, our results are important steps towards automatic systems able to analyze, assist and modify small group dynamics in order to provide various kinds of support to dysfunctional teams, from facilitation to training sessions addressing both the individuals and the group as a whole. On a more theoretical side, our work emphasizes the relevance of micro-level interaction dynamics in determining if the group performance will be good or bad. Of course, more work is needed to fully explore the power of the model proposed and to validate our findings using for example larger samples of data and analyzing different kinds of small group interactions (brainstorming, competitive games, etc.).

REFERENCES

[1] Boys, R. J., Wilkinson, D. J. and Kirkwood, T. B. L. Bayesian inference for a discretely observed stochastic kinetic model. *Statistics and Computing*. 18(2) 125-135.

[2] Hall, J. W. and Watson, W. H. 1970. The Effects of a Normative Intervention on Group Decision-Making Performance. *Human Relations*. 23(4) 299-317.

[3] Harrigan, J.A. Listeners' Body Movements and Speaking Turns. 1985. *Communication Research* 12(2) 233-250.

[4] Karau, S.J, and Williams, K.D. 1993. Social loafing: A meta-analytic review and theoretical integration. *Journal of Personality and Social Psychology*. 65 681–706.

[5] Kendon, A. (1967). Some Functions of Gaze Direction in Social Interaction. *Acta Psychologica*. 26 22–63.

[6] Kerr, N.L. and Tindale, R.S. 2004. Group performance and decision making. *Annual Review of Psychology*. 55 623-55.

[7] Kerr, N. L., & Kaufman-Gilliland, C. M. 1994. Communication, commitment, and cooperation in social dilemmas. *Journal of Personality and Social Psychology*. 66 513–529.

[8] McGrath, J. E. 1984. Groups: Interaction and Performance. Prentice-Hall, Englewood Cliffs, NJ.

[9] Olguín, D., Waber, B., Kim, T., Mohan, A., Ara, K. and Pentland, P. 2009. Sensible Organizations: Technology and Methodology for Automatically Measuring Organizational Behavior. *IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics*. Vol. 39, No. 1, pp. 43-55.

[10] Opper. M. and Sanguinetti. G. 2007. Variational inference for Markov jump processes. *NIPS 2007*.

[11] Sacks, H., Schegloff, E. A. and Jefferson, G. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language*. 50(4) 696-735.

[12] Shaw, M. E. 1976. Group dynamics: the psychology of small group behavior. New York, McGraw-Hill.

[13] Wilson, D. S., Timmel, J. J. and Miller, R. R. Cognitive cooperation: When the going gets tough, think as a group. *Human Nature*, 15(3):1–15, 2004.

[14] Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N. and Malone, T. W. 2010. Evidence for a collective intelligence factor in the performance of human groups. *Science*. 330(6004) 686-68

[15] Yngve, V. On getting a word in edgewise. 1970. *Papers from the sixth regional meeting of the Chicago Linguistic Society*, pages 567–77, Chicago: Chicago Linguistic Society.