Quantifying Group Problem Solving with Stochastic Analysis

Wen Dong, Alex "Sandy" Pentland
MIT Media Laboratory
(wdong,sandy)@media.mit.edu

ABSTRACT

Quantifying the relationship between group dynamics and group performance is a key issue of increasing group performance. In this paper, we will discuss how group performance is related to several heuristics about group dynamics in performing several typical tasks. We will also give our novel stochastic modeling in learning the structure of group dynamics. Our performance estimators account for between 40 and 60% of the variance across range of group problem solving tasks.

Keywords

Index Terms—Human dynamics, speech communication, interaction process analysis, stochastic modeling

1. INTRODUCTION

We are interested in quantifying group problem solving performance by analysis of non-linguistic social signals. These pre-linguistic communication structures have been shown to capture a large fraction of the dynamics of group interaction, and to be predictive of performance and outcomes [1, 2]. We accomplish this by instrumenting group participants using *Sociometric Badges* [3, 2], to record speaking dynamics, tone of voice, body motion, etc. These data are then analyzed by use of signal processing techniques including HMM, influence, and similar stochastic models.

The *Interaction Process Analysis (IPA)* [4] is a traditional approach for quantifying a general group problem-solving process based on fine time-grained analysis. In this approach, an interaction process is treated as a sequence of events of different categories — giving and analyzing facts, showing individual approaches for problem-solving, making group decisions, and releasing tensions developed in decision-making. The analysis proceeds in the following way: Two or more trained observers watch through a whole group problem-solving process and mark events at a resolution of 10˜15 events per minute; The sequences of events marked by different observers are then compared and accessed for reliability; Heuristic scores about the interaction process are then computed by counting events in different categories and are related to group performance.

While they could quantitatively explain the relationship between the details of an interaction process and the corresponding group performance, the traditional methods are costly in terms of human expert time. As a result, there are many difficulties in applying these methods in explaining the fine differences about the interaction dynamics and performances of a large number of groups in solving a large number of different problems.

On the other hand, we argue that the traditional approaches could be complemented, automated and unified with a new approach based on the statistical learning methods and our capability to collect a massive amount of data about group interaction processes with embedded devices. Our reasoning is the following. Different types of activities in a group problem-solving process have different temporal and interaction statistics — Fact-giving often involves longer sentences and less parallel-speaking from other speakers while showing-opinions often involves shorter sentences and more parallel-speaking. Further, the solutions of many common problems often involve a limited amount of facts, opinions and voting and thus a limited amount of events of different categories in problem-specific proportions. Thus we could estimate group performance based on heuristics and stochastic methods about these non-semantic cues of the group process, and potentially find ways to improve it. In situations when we do not know the structure of the group problem-solving process, we could use latent-state stochastic models to "project" the time series of non-semantic cues along the direction of problem-solving performance and discover the structure of problem solving.

To illustrate our method in quantifying group interaction-dynamics and problem-solving, we will refer to the interaction-dynamics data collected by the Sociometric Badges in the *Measuring Collective Intelligence (MCI) study 1* [5]. The goal of the MCI studies is towards finding the key components of collective intelligence, the relationship between group interaction and group performance, and the methods to increase collective intelligence. *MCI Study 1* involves 42 groups solving 12 problems, with each problem costing a fixed amount of time ranging from 10 minutes to 1 hour and all 12 problems costing around 3 hours.

In the rest of the paper, we will summarize some key feature extraction steps, discuss some heuristics about quantifying group problem solving, and give our stochastic model-

ICMI-MLMI'10, November 8-12, 2010, Beijing, China.

ing that learns the structure of group problem solving.

2. DATA PREPROCESSING

Throughout the MCI studies, we instructed each subject to wear a sociometric badge through a lanyard around the neck. Each badge recorded the audio of its wearer, the movements of its wearer (through the accelerometer), the orientation of the wearer relative to other participating group members (through the infrared interface), a sequence of button-presses (by the wearer at the task boundaries according the instructions to press the single button on the badge and used by us to mark the beginnings of tasks), and a periodic sequence of messages from the other badges that contained the senders' local times and the receiver's local time (through the BlueTooth interface). The message sequences were used to align the signals from different badges in the MCI studies.

Since we are interested in comparing the interactions and the performances of different groups in solving different problems, we translated the button-presses into the task boundaries by finding the Viterbi path of a hidden Markov model in which the observations were the time-intervals between neighboring button-presses and the latent states were the transitions from task boundaries to later task boundaries. The parameters of the hidden Markov model were set according to the manually marked task boundaries for three groups that we chose.

We aligned the local times of different badges used in a same group process through the principle component analysis (PCA) of the messages that contained the senders' local times and the receivers' local times. We subsequently took the first principle component as the global time and used the relationships between local-times and the global time to adjust the times of other time series. When the audio recordings contained speaking, we also aligned the local times of different badges by aligning the pitched segments recorded by the badges. Due to their duration and spacing statistics, the pitched segments in different badges could in most cases be unambiguously aligned (Fig. 1).

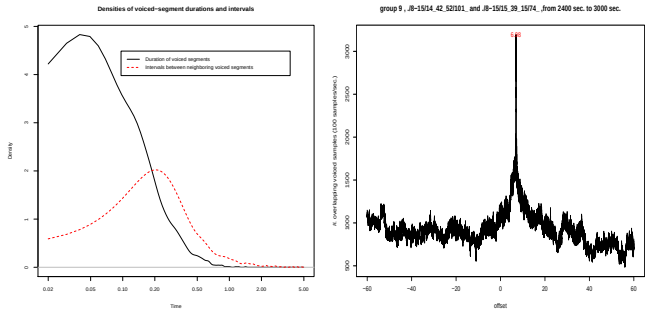

Figure 1: Left: The voice segments in our data set normally last 0.05 second and no longer than 0.5 second; They are normally 0.2 seconds apart when appearing in the same clause. Right: Aligning voiced frames could be a robust way to align data collected by different embedded devices deployed in close distance; When aligned, the pitch signals collected by different embedded devices are normally equal to each other.

We extracted the voiced segments using the 9-parameter algorithm of Boersma [6], which was reported to have small pitch determination error and large resolution of determination of harmonics-to-noise ratio due to its method of computing the short-term autocorrelation function of continuous-time audio time series. We estimated who is speaking by comparing the sound intensities of the voiced segments recorded by different badges. Based on our investigation of a random sample of 10 minutes recordings of different groups, the voiced segment detection algorithm could archive about 95% precision, 90% recall, and the speaker detection algorithm could achieve about 95% accuracy. The speaking and speaker features, together with other features of the interaction processes, were imported through some scripting languages into Audacity as label tracks and into Praat as TextGrid for investigation.

3. PERFORMANCE HEURISTICS

In this section, we will discuss several statistics about the group problem-solving processes that are related to group performances.

We will first show the relationship between the total number of clauses in an interaction process and the corresponding performance in the cases of an easy brainstorming task, a task to solve analytical IQ problems, and two tasks about optimization with constraints. This type of relationship is exploited by people to evaluate the scale of a piece of software based on its lines of source code, to evaluate the difficulty of a problem set based on the number of pages need to write down the full solutions, and to evaluate the proficiency of a person based on the number of work pieces he could finish in unit time.

In the MCI Study data sets, most of the discourses in the interaction processes are directly related to the solution of the problems, since most groups took their tasks seriously in most of time. The groups spoke the same contents in the same dynamics to solve the problems up to rephrasing, sentence permutation and the addition of some supplementary sentences, since the groups were required to solve the problems together by communication and there was normally one way to solve each problem. Hence the performances of the groups were not only determined by what the group members said but also significantly correlated with how the group members spoke. We could not only estimate group performance of such a task by counting the number of clauses in the interaction process but also give a prescription for improving group performance based on factors such as average clause length and speaking speed.

Performance Heuristics

We count the number of clauses in an interaction process using a hidden Markov process. The hidden Markov process has two latent states. Corresponding to each voiced segment, the observation of the hidden Markov process is comprised of whether there is a speaker change and the time interval between the current and the past voiced segment, with the two elements of the observation independent of each other. After it is fitted with interaction processes, such a hidden Markov process normally contains one latent state corresponding to no speaker-change and a sub-second interval (approximately 0.2 second) between neighboring voiced segments, and another latent state corresponding to a significant probability of speaker-change (normally greater than 30%) and an interval of more than one second. In the rest of the section, we will use the latent state with longer time

$$\begin{aligned}
\text{brainstorming score} &= \underset{R^2=0.62,p=0.01}{(\text{number of clauses} - 127)/1.4} \\
\text{group IQ score} &= \underset{R^2=0.6,p=0.006}{(\text{number of clauses} - 67)/11} \\
\text{making judgment score} &= \underset{R^2=0.62,p=0.13}{\text{number of clauses}/4} \\
\text{shopping task score} &= \underset{R^2=0.37,p=0.08}{(\text{number of clauses} - 219)/1.8}
\end{aligned}$$

Table 1: Each task in the MCI data set involves a similar preparation stage and a series of similar facts to be collected, hence we can estimate performance by counting the number of clauses in a group interaction process. The p-value for making-judgment task is not significant because many data points are unusable.

interval as the indicator of the start of a clause, and subsequently compute the number of clauses, as well as other statistics, of an interaction process.

The overall activity level in a group problem solving process gives us information to estimate the overall group performance, and the activity level at a finer time resolution gives us information to track the performance over time (Table 1). In the MCI data set, there exist strong linear relationships between the number of clauses and performance score in group brainstorming and group IQ test. The number of clauses and the performance score in the processes of the judgment task and the shopping task, on the other hand, may not have linear relationships while they are positively correlated. In order to figure out the relationship in the judgment task and the shopping task, we need either an understanding of how people really solve the two tasks or a larger sample of the interaction processes for solving the two tasks.

4. LEARNING GROUP PROBLEM-SOLVING STRUCTURE WITH STOCHASTIC MODELING

While it is a good step forward to quantify the interaction processes using heuristics based on signals recorded by embedded devices and to explain performances thereby, the heuristic-based approach has several limitations. Firstly, the approach still costs a good amount of expert time to figure out the right heuristics for different types of tasks, and sometimes the statistics that differentiate good and bad performances could be complex and delicate. Secondly, the approach is sensitive to and does not discriminate outliers, such as when a group did not work on what it was supposed to do. As a result, we will discuss in this section a non-parametric approach of learning the structure of group problem solving, that uses mixture of hidden Markov processes (HMPs) modeling to describe the probability measure of an interaction process, with each component HMP in charge of explaining the dynamics-performance relationship of solving one of the four specific types of problems.

There are similarities between the mixture of HMPs modeling and how humans figure out the dynamics-performance relationships about group problem solving. Given a training set of group interaction processes of problem solving, labeled with their problem types and performance scores, a human observer will intuitively relate the interaction processes to the corresponding problem types and performance scores. He will assign meanings to different parts of the processes, and tell the differences among the dynamics related to different problems and different performance scores in terms of some statistics such as average clause length, speaking speed, and the frequency of transitions to different speakers. Given the observation of a new interaction process, he will compare the new process with the template processes in the training set, and tell (a) whether the process was intended to solve any problem, (b) which type of problems the process was intended to solve most likely, and (c) which (latent) performance covariate could best explain the dynamics in the process.

In the mixture of HMP model, any sequence $(S_t, O_t)_{t=0\cdots T}$ of latent-state and observation tuples is sampled with probability w_i from hidden Markov process i out of the $n+1$ different hidden Markov processes parameterized by θ_i where $0 \leq i \leq n$. Thus $\mathbb{P}\left((S_t, O_t)_{t=1\cdots T}\right) = \sum_{i=0}^{n} w_i \mathbb{P}\left((S_t, O_t)_{t=1\cdots T}; \theta_i\right)$. Of the $n+1$ hidden Markov processes, process θ_0 is the garbage process that explains everything else, and process $1 \leq i \leq n$ explains the dynamics of the interaction processes in solving task i. The parameters (i.e., the state transition matrix, and the parameters related to the observation model) $\theta_i = \{A_i, B_i\}$ for processes $1 \leq i \leq n$ are functions of the performance covariate f and the parameters θ_0 are constant. We take linear functions in our modeling: $A_i(f) = A_i^{(0)} + f \cdot \alpha_i$ and $B_i(f) = B_i^{(0)} + f \cdot \beta_i$ with the constraint that $A_i(f)$ and $B_i(f)$ are valid. With the given definition, model fitting follows the standard EM algorithm. After it is fitted interaction processes for different tasks, the model is used for finding the most likely mixture component and the performance covariate $\text{argmax}_{i,f} \mathbb{P}\left((S_t, O_t)_{t=1\cdots T}; f, \theta_i\right)$ for given $(S_t, O_t)_{t=1\cdots T}$.

Performance

From the 43 interaction processes in solving the four tasks, we made 1000 draws of testing sets of 8 processes each, and corresponding to each test set we used the rest 35 processes to train the mixture model. Overall we could estimate the performance of the processes in the testing sets with $R^2 = 60\%$ accuracy ($p < 0.01$) by using the mixture model trained with the corresponding training sets and taking the maximum likelihood performance covariate.

The performance coefficients of the fitted models for the four tasks (Table 2) tell us not only how different tasks require different group process dynamics but also how different performances in the same task correspond to slightly different dynamics. Good performance generally requires active discussions (e.g., the coefficients in the first four rows are generally positive). On the other hand, the brainstorming task and the group IQ task both have faster speaker transitions ($\mathbb{P}(\text{chg.spkr}|s_1)$), shorter clauses ($\mu(\Delta t|s_1)$), and longer pauses ($\mu(\Delta t|s_2)$) than the group shopping task and the group judgement task. Further, better performances in the brainstorming task and the group IQ task normally requires faster speaker changes, longer clause lengths and less standard deviations of pauses ($\sigma(\Delta t|s_2)$). The different dynamics in the two types of tasks are due to the fact that brainstorming and IQ problems normally requires a good aptitude of making discoveries through unusual paths, while a planning a shopping itinerary and making a judgement normally

	b.s.	grp.iq	shop	jdgmnt
$\mathbb{P}(s_1 \rightarrow s_1)$	.9-f*2e-5	.9-f*3e-5	.9	.9
$\mathbb{P}(s_2 \rightarrow s_2)$	.2+f*2e-5	.1+f*2e-4	.2	.2
$\mathbb{P}(\text{chg.spkr}\|s_1)$	.2+f*1e-5	.3+f*1e-5	.1+f*1e-5	.2
$\mathbb{P}(\text{chg.spkr}\|s_2)$	.4+f*1e-5	-.03+f*4e-4	.5	.4
$\mu(\Delta t\|s_1)$	.2+f*1e-5	.2+f*8e-5	.3	.2
$\sigma(\Delta t\|s_1)$	.1+f*2e-5	.1+f*9e-5	.2	.2
$\mu(\Delta t\|s_2)$	2.0+f*7e-5	3-f*3e-4	1.7	2
$\sigma(\Delta t\|s_2)$	3.2-f*2e-4	7.0-f*2e-1	2.7+f*2e-5	9

Table 2: The HMP parameters for the four tasks summarize the different dynamics-performance relationships in both the task dimension and the performance dimension. In this table, covariate f represents performance score, latent state s_1 and s_2 respectively represent the state of making progress and the state of not making progress, and the observations are change of speaker and duration of clause/silence.

involves making good reasoning. The longer clause lengths in brainstorming and solving IQ problems correspond to actively giving information rather than passively accepting an answer, and the less standard deviations of pauses correspond to consistent performance throughout a task.

We can proceed to simulate the fitted component HMMs at different performance levels, and sample the performance heuristics corresponding to different tasks (Table 3). The heuristics agree with our observation (Sec. 3) about the positive correlation between active discussion and good performance, as well as the different dynamics required by solving different tasks. For one example, at the 25%, median, and 75% performance levels, the interaction processes to solve the group IQ problem will respectively produce 10.6, 12.3, 14 clauses and involve 4, 5, 7 speaker changes per minute. For another example, on average the fractions of clauses longer than 4 words in planning shopping itinerary and making judgment are noticeably longer than the fractions in brainstorming and solving group IQ problems.

We are interested in quantitatively reasoning about the different approaches to improve group performance through our modeling. Effectively mixing the ideas of the group members (e.g., through encouraging faster speaker turns and shorter sentences) generally helps improving performance. Hence we would encourage group members to actively contribute to problem solving but not to manipulate it. Groups with longer speaking turns in the MCI brainstorming and Group IQ tasks do not necessarily have better scores in these two tasks. Hence if these dynamics-performance relationships in the MCI data are typical we would encourage future groups to have shorter turns in similar tasks such as the MCI brainstorming and group IQ tasks.

5. CONCLUSIONS AND DISCUSSIONS

In this paper, we discussed our approach to make embedded devices understand group problem-solving. We do so by relating the performance score with how the group solves the problem for any group problem-solving process. Specifically, by using statistics such as number of clauses, number of vowels, speaking speed, clause length, and cycles of serial speaking and parallel speaking, we could estimate group performance score with the R-squared value up to 40%. By stochastic modeling of the group problem-solving process that is conditioned by the type of task and the performance score, we could learn the structure of group problem-solving and achieve higher accuracy in estimating performance score. Our code is available at `http://vismod.media.mit.edu/vismod/demos/influence-model/index.html`

percentile	number of clauses	clauses per person*minute	speaker. turns per minute	vowels per clause	speaker overlap
25%	250	10.6	4	2	0.8
50%	300	12.3	5	1.5	1.2
75%	350	14.0	7	1.2	1.4

(a) This table is constructed by sampling several performance statistics at different performance levels from the trained hidden Markov model of solving the MCI brainstorming task. It can be used to look up the MCI brainstorming performance.

task	1 word	2 words	3 words	>3 words
brainstorming	33%	19%	12%	36%
group IQ	36%	18%	11%	35%
itinerary planning	32%	17%	10%	41%
making judgment	32%	16%	10%	42%

(b) This table is constructed by sampling clause lengths at the median performance level from the trained hidden Markov models of all four MCI tasks. It captures the fact that the MCI brainstorming and group IQ tasks involve more clauses that are 4 words or less.

Table 3: A stochastic model of group problem solving enables us to explore the different heuristics of estimating performance without conducting further expensive experiments.

6. REFERENCES

[1] Wen Dong, Taemie Kim, and Alex Pentland. A quantitative analysis of collective creativity in playing 20-questions games. In *Proceedings of the ACM Conference on Creativity and Cognition*, 2009.

[2] Alex Pentland. *Honest Signals: how they shape our world*. MIT Press, 2008.

[3] Daniel Olguin Olguin, Banjamin N. Waber, Taemie Kim, Akshay Mohan, Koji Ara, and Alex Pentland. Sensible organizations: Technology and methodology for automatically measuring organizational behavior. *IEEE Transactions on Systems, Man, and Cybernetics — Part B: Cybernetics*, 39(1), 2009.

[4] Robert Freed Bales. *Interaction Process Analysis: a Method for the Study of Small Groups*. Addison-Wesley Press, 1950.

[5] Measuring Collective Intelligence Studies. `http://cci.mit.edu/research/measuring.html`.

[6] Paul Boersma. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. In *Proceedings of the Institute of Phonetic Sciences*, volume 17, pages 97–110, 1993.