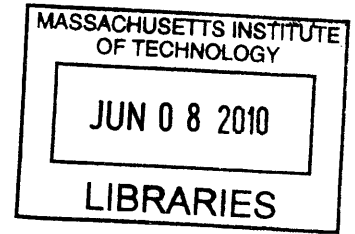# Modeling the Structure of Collective Intelligence

by

Wen Dong

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
at the
Massachusetts Institute of Technology

**ARCHIVES**

June 2010

Signature of Author .................................................................
Program in Media Arts and Sciences
May, 2010

Certified by ...........................................................
Alex (Sandy) Pentland
Toshiba Professor of Media Arts and Sciences
Thesis Supervisor

Accepted by ...........................................
Pattie Maes
Chair, Program in Media Arts and Sciences

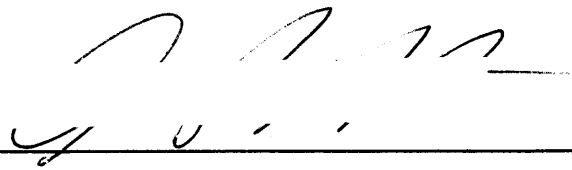# Modeling the Structure of Collective Intelligence

by

Wen Dong

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
at the
Massachusetts Institute of Technology

June 2010

Thesis Advisor                                         Alex (Sandy) Pentland
Professor of Media Arts and Sciences
Massachusetts Institute of Technology

Thesis Reader                                          David Lazer
Associate Professor of Public Policy, Kennedy School of Government
Harvard University

Thesis Reader                                          Henry Lieberman
Research Scientist, MIT Media Laboratory
Massachusetts Institute of Technology

# Modeling the Structure of Collective Intelligence

by
Wen Dong

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning
on June, 2010
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

## Abstract

The human problem solution process has attracted an increasing amount of interest among educators, managers, computer scientists and others. However, the discussion of the subject has suffered from the lack of stochastic tools to quantitatively capture both the subtler steps of the problem solution process and the diversity of human thinking. In order to stochastically model the human problem solution, this thesis presents an approach referred to as "influence modeling," that attempts to describe how an individual navigates from one random memory chunk to another related memory chunk, and how a group of people randomly remind one another of memory chunks that could be individually uncommon.

As an application of influence modeling, this thesis shows how groups play "20-questions" games based on a semantic space, such as ConceptNet (a common-sense database). It also investigates how groups send signals about their behavior, which are collected by embedded devices, how group interaction processes could be automatically monitored with embedded devices, how group performance could be facilitated, and how to map group behavior and performance from the macroscopic level to the microscopic level in experiments in measuring collective intelligence. The influence modeling makes it possible to understand how a group could perform better than an individual. It also allows for the monitoring of the status of the problem solution, and makes it possible to direct group interaction in more fruitful ways.

Thesis Advisor: Prof. Alex (Sandy) Pentland
Toshiba Professor of Media Arts and Sciences

# Contents

# Chapter 1

# Introduction

My research aims to help individuals and groups to improve their problem-solving ability. I approach this goal by modeling how people solve real-world problems. From the perspective of a computer scientist, I emphasize modeling the mechanism of individual and group problem solving. I base this modelling process upon stochastic methods and signals that can be easily collected with embedded devices. After this, I then use these models to help guide individuals and groups toward more productive behaviors.

I have developed a method I refer to as "influence modeling" to measure, describe and understand the momentary roles people take during a group problem solving process, and to quantify the performance of the group based upon these role-taking statistics. These influence modeling procedures enable us to detect rare and important roles related to raising of disagreements and the asking of questions with approximately 50% accuracy and recall. This method also allow for the detection of common and less important roles related to the giving of facts and passive listening with approximately 85% accuracy, and to estimate performance from behavior with $R^2 \approx .6$.

Based upon this modeling approach, I have also been able to model and confirm several psychological hypotheses. For instance, I have been able to demonstrate that active conflict-resolution among group members increases group problem-solving performance. This is based on a widely-accepted management principle, and observations of the different socio-emotional roles and task roles taken by the group members. This has especially been true for the roles related to conflict resolution.\

Let us first consider the concept of describing human problem-solving (including learning) as a searching process in a network of human knowledge, or what might be referred to in more formal terms as a derived network. Since people have different (but comparable) backgrounds and thinking patterns, a sample of individual/group problem-solving processes collected should reveal both commonalities and individuality in terms of the strategies and performances of different people.

The key mathematical tool I developed for modeling the problem-solving dynamics of individuals and groups of different sizes is called the latent state influence model. This tool captures the group dynamics,as demonstrated in the following brainstorming task. In this task, a group is asked to generate as many ideas as possible in a specified amount of time that satisfy a given set of constraints.

Building on several existing psychological theories,, I model the set of ideas the people generate, as well asthe set of dyadic relations of each other individual within the network. In this network, the individual can navigate from one idea to another via a relationship, with the rate at which they move being dependent upon both the individual and their relationship with the other people. Hence, the brainstorming process could be modeled with a set of "reactions" $w_i \to w_j$ from idea $w_i$ to idea $w_j$ with a corresponding transition rate $p^{(c)}_{w_i \to w_j}$ that is dependent on individual $c$.

In experimental scenarios we observe "social signals," such as interruption and turn taking, that only give incomplete information about the transitions. Latent state models can be used to determine which signals are more indicative of transitions. Since the goal is to help people in their problem solving, we are interested in the domain knowledge structure.It is also important to observe human behavior and group interaction related to general intelligence (such as intention and memory). The factors that are orthogonal to the direction of problem-solving performance can be ignored.

A discrete-state stochastic modeling of human activities normally involves a large number of latent states related to what people think, and, correspondingly, a complex state-transition kernel. Influence modeling simplifies the state-transition kernel by imposing further structure upon the state transitions. It is one of many tools that can be used to investigate behavior related to problem solving in a laboratory setting.

In a discrete time hidden Markov process with finite latent states $\{1, \cdots, N\}$, state transition events $i \to j$ happen at discrete times $t \in \{1, 2, \cdots\} = \mathbb{N}$. The events happen with probability $\mathbb{P}(S_{t+1} = j | S_t = i) = a_{ij}$. These observations contain partial information about the latent states based on a probability measure $\mathbb{P}(O_t | S_t)$. In comparison, in a latent structure influence process, a system has finite components (chains) $\{1, \cdots, C\}$, and each component has finite latent states $\{1, \cdots, N_c\}$. The observations for each component contain partial information about the latent states for the same component based on probability measures $\mathbb{P}(O^{(c)}_t | S^{(c)}_t)$. In this influence process, components cause each other to change states $S^{(c')}_t = i \rightsquigarrow S^{(c)}_{t+1} = j$ at discrete times with probability $\mathbb{P}(S^{(c')}_t = i \rightsquigarrow S^{(c)}_{t+1} = j) = h^{(c',c)}_{i,j}$. Each component state change has one exact cause. The discrete-time hidden Markov model and discrete-time influence modeling have their (continuous-time) jump Markov model counterparts. Hence, the influence modeling captures the interactions among different components of a system.

My research seeks to understand human problem solving by studying how stochastic processes that are compatible with psychological findings search through networks related to human knowledge and find answers. If the models that describe important areas of human problem solving (such as reading, writing and creativity) have good predictive power, then they could potentially be very useful in practical ways. people,

In addition to providing mathematical support for the development of new human-computer interaction interfaces, I am interested in helping organizations and societies relax from information overload, regain self-awareness, and focus attention upon their real needs. We may eventually see the emergence of a new computational social science based upon stochastic modeling, that aims to create a lucid society, provides privacy and offers help of the type that is mostly urgently needed.

I am also interested in expanding my work from the domain of the small group behavior and performance into investigation of the whole society, again using a data-driven approach. For example,

finding the structure in the publication of a whole research field — who influences who, the importance and key contributions of each article in different levels of detail, and what are the current hot directions — could be an interesting research project.My experiences in finding the local knowledge structures and modeling the problem-solving processes from small-group problem-solving dynamics and organizational dynamics could be very useful in such an endevour..

Just as stochastic methods can be used to model human behavior, it is also possible to use human problem-solving to inspire mathematical optimization and solution techniques. From my modeling of human problem solving as a diffusion process, I can proceed to compare how humans and computer algorithms solve similar problems and to find interesting theorems. Influenced by the approach of mathematical logic, I believe the meaning of an entity exists in its relationship with other entities. Hence, it is reasonable to abstractly discuss the strategies and the corresponding observable behavior of solving, for instance, a 20-questions game. It is also reasonable to believe the observable behavior and strategies involved in solving a 20-questions game will also be applicable to those used for solving a wide variety of other problems.

# Chapter 2

# Diffusion Model of Human Problem Solving

I use the following influence process to capture how people influence one another and accordingly change their thinking. The influence process is a Markov chain computationally much cheaper approximation of the continuous-time Markov jump process to study the dynamics of complex networks.. The influence process that I developed here was inspired by the work of Asavathiratham [Asavathiratham, 1996].

In a imaginary world such as the network of flipflops (Figure 2.1), each flipflop can take two states: red and blue, each flipflop changes its state occasionally by itself or by an occasional message from another flipflop that takes an opposite state, and each flipflop occasionally misrepresents itself as having an opposite state. We are often interested the following properties about this network. How are the flipflops connected with each other, for in real-world the relationships among the "flipflops" may not be obvious and even in this visually intuitive imaginary world the network will behave differently according to the strengths of the connections? How does the network evolve over time? Who are the most influential flipflops, for we surely want to target the most influential flipflops if we want to run a campaign on the network? How well can the network solve problems and how much performance variation can the network have if the performance of the network can be expressed as an integral of a function of the flipflops' states? We often claim intuitively that collaboration increases the convergence rate and decreases the diversity in problem solving, and the network dynamics model may provide us a tool to quantitatively reason about our intuition.

Specifically, I hope to capture the relationship between information flow and performance at both macroscopic social network level and microscopic knowledge network level. At the social network level, I am interested in how critical information could be quickly spread in the network through the connectors and the persuaders to make things happen [Gladwell, 2002]. At the microscopic level, I am interested in how individuals and groups could quickly fetch the critical pieces of knowledge and have problems solved.

In a real world network, we estimate the structure of the network and the states of the nodes by counting the concurrences of the states of the nodes. We often need to iterate the structure learning and the state estimation processes for many times to get a good understanding of the network. The more knowledge we have about the structure of the network, the more precision we can get in estimating the states of the nodes, and vice versa. The statistical algorithms described here learn the structure of the flipflop network in the same way that we estimate the real world network behind

all the maths (Figure 2.2). We can tell the misrepresentations of the flipflops' states by looking at their previous and later states as well as their neighbours' states, since the misrepresentations are occasional and are not systematically organized with one another. We can tell the (unknown) network structure and how the state of one flipflop influence the state of a neighbouring flipflop by noticing how often a pair of states appear one after another, no matter whether a blue node makes a red neighbouring node, a blue node makes a blue node, or otherwise. When two nodes go further and further from each other, the correlation between their states becomes weaker and weaker.

The preceeding statisitical inference about the flipflop network assumes both that we can observe all events that changes the flipflops' states and that we cannot directly observe the flipflops' states. These assumptions are similarly taken in many other latent state models used in the areas of digital signal processing (DSP). The first assumption reflects the optimism of the DSP researchers that they can develop devices fast enough to capture almost all events of their interest. This assumption is generally not taken by computational systems biologists in studying the change of chemical concentrations (latent states) caused by chemical reactions (events). Hence computational systems biologists have to either develop intriguing Metroplis-Hastings scheme to sample the possible event sequences that lead from one chemical concentration the next, or use approximations and other assumptions to work around the need to estimate the event sequences [Wilkinson, 2006]. The second assumption is due to the diversity and randomness of human behavior. For instance, a person may not know that he possesses a critical resource or certain opinions and will report imprecisely about those "states". Often we only have limited knowledge to map our observations about people to the real states.

In the rest of this section, I will discuss what the influence model is, and how we can make statistical estimations with the influence model.

## 2.1   Influence Model

Let us consider how a node in a graph changes its state ($S_t^{(c)}$ where $c$ is the node index and $t$ is the time index) over time as the result of the occasional and independent influences sent from one another nodes along the edges ($H^{(c',c)}_{S_t^{(c')},S_{t+1}^{(c)}}$). Let us further put an additional constraint/approximation that node $c'$ has constant overall influence rate on node $c$:

$$\forall i,j : \sum_{S_{t+1}^{(c)}} H^{(c',c)}_{S_t^{(c')}=i,S_{t+1}^{(c)}} = \sum_{S_{t+1}^{(c)}} H^{(c',c)}_{S_t^{(c')}=j,S_{t+1}^{(c)}},$$

$$\Leftrightarrow H^{(c',c)}_{S_t^{(c')},S_{t+1}^{(c)}} = D^{(c',c)} A^{(c',c)}_{S_t^{(c')},S_{t+1}^{(c)}} \text{ where } \sum_{S_{t+1}^{(c)}} A^{(c',c)}_{S_t^{(c')}=i,S_{t+1}^{(c)}} = 1.$$

At any time $t$, a node $c$ currently in state $S_t^{(c)}$ receives a message from a node $c'$ that asks node $c$ to change state to $S_{t+1}^{(c)} \neq S_t^{(c)}$ at a rate/probability $H^{(c',c)}_{S_t^{(c')},S_{t+1}^{(c)} \neq S_t^{(c)}}$, and the overall rate/probability of
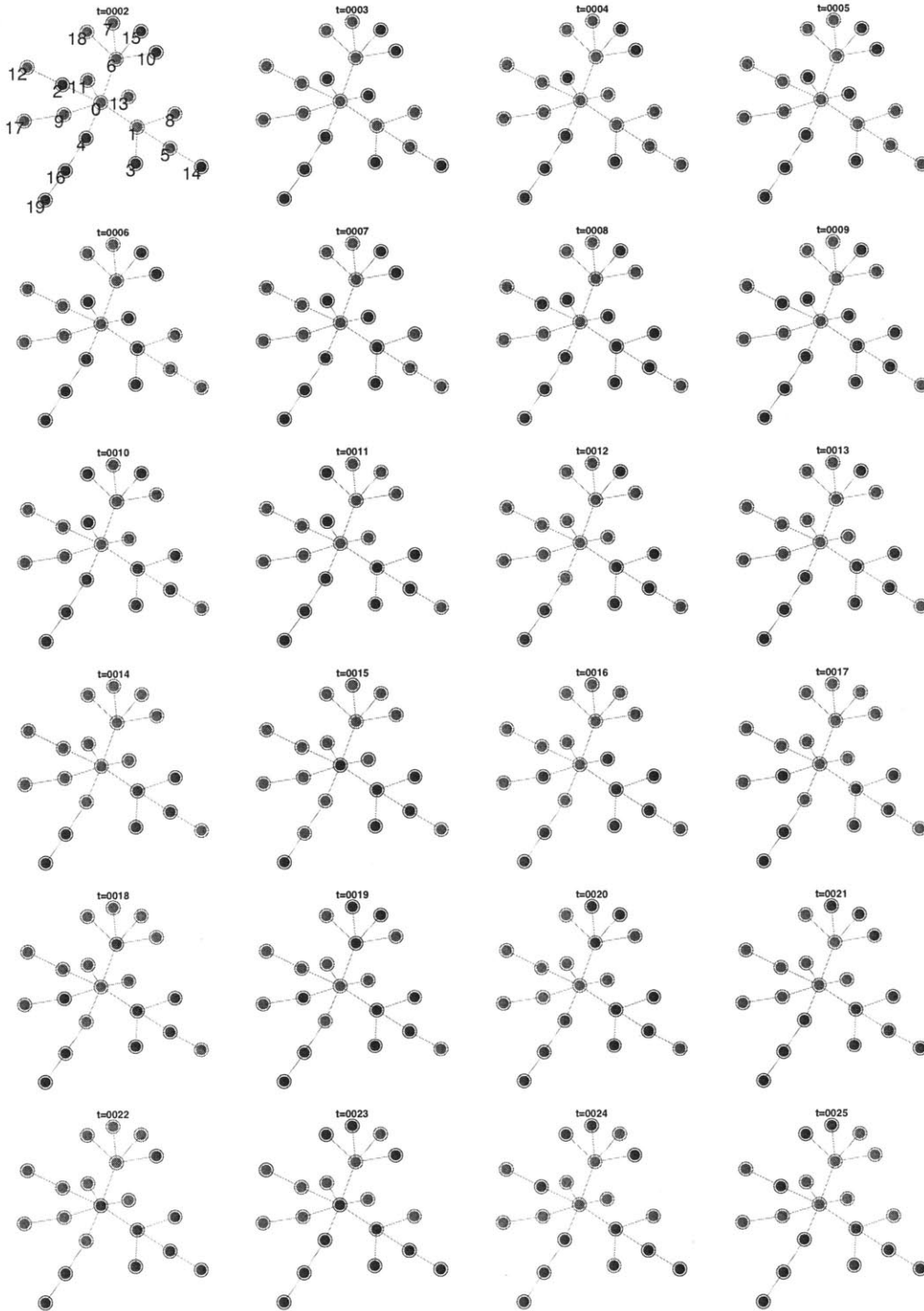
Figure 2.1: A network of flipflops.
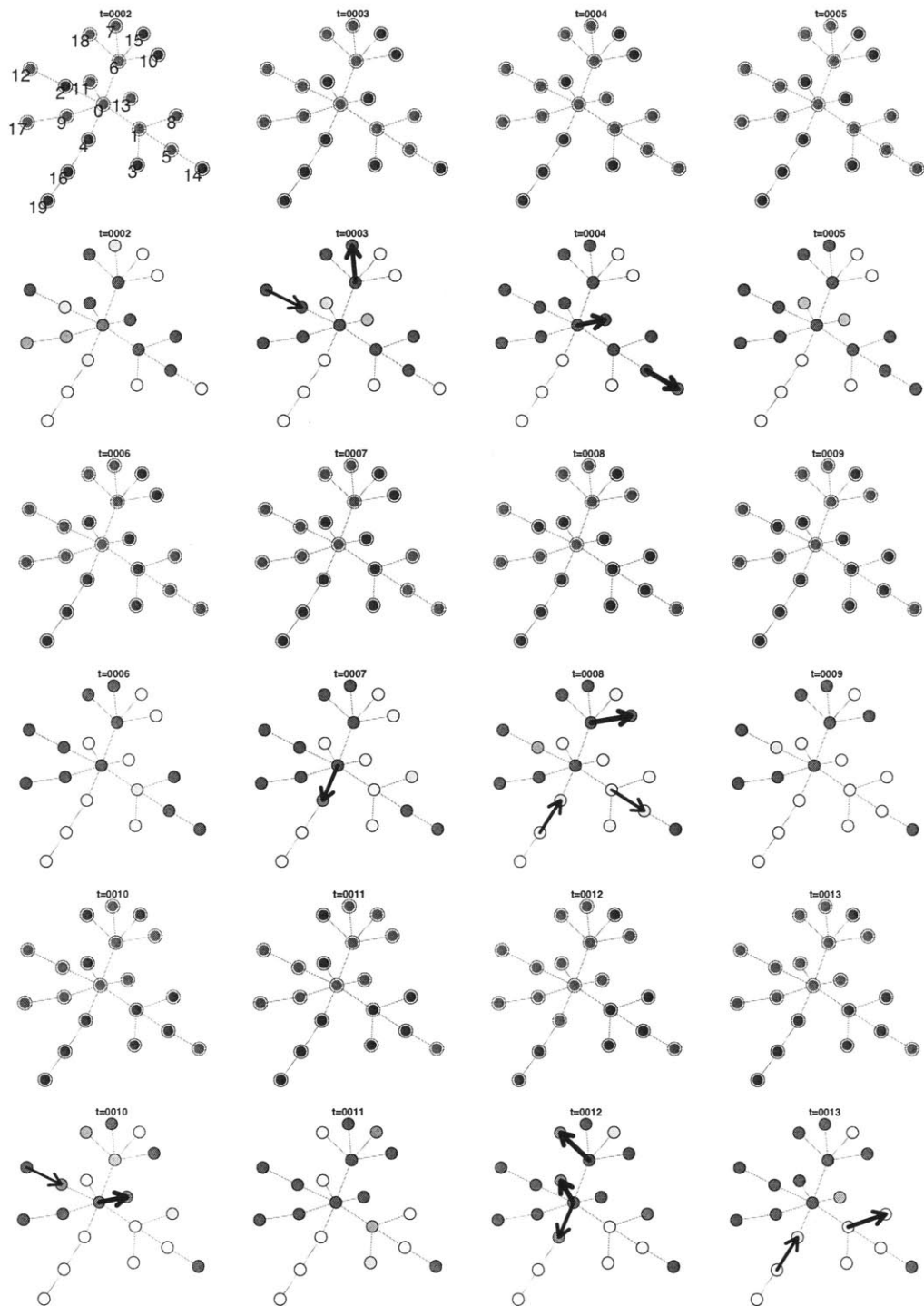
Figure 2.2: A network of flipflops (learning).

existing such a message and a state change to $S_{t+1}^{(c)}$ is

$$\mathbb{P}\left(S_{t+1}^{(c)}|S_t^{(1)}\cdots S_t^{(C)}\right), \text{ where } S_{t+1}^{(c)} \neq S_t^{(c)}$$

$$= \mathbb{P}\left(\exists \text{ a message that requests node } c \text{ to go to } S_{t+1}^{(c)} \neq S_t^{(c)}\right)$$

$$\overset{\text{inclusion-exclusion principle}}{=} \sum_{c'}\mathbb{P}\left(\text{message from node } c'\right) - \sum_{c'\neq c''}\left(\text{message from } c' \text{ and } c''\right) + \underset{c',c'',c''' \text{ mutually different}}{\sum} \cdots$$

$$\overset{\text{probability that more than 2 messages received is negligible}}{\approx} \sum_{c'} H_{S_t^{(c')},S_{t+1}^{(c)}}^{(c',c)} \quad .$$

The probability that node $c$ stays in state $S_t^{(c)}$ is hence $1-\sum_{S_{t+1}^{(c)}\neq S_t^{(c)}}\sum_{c'}H_{S_t^{(c')},S_{t+1}^{(c)}}^{(c',c)} = \sum_{c'}H_{S_t^{(c')},S_t^{(c)}}^{(c',c)}$, which takes the same form as the probability that node $c$ changes state. The overall probability that the whole graph system goes from state $S_t^{(1)}\cdots S_t^{(C)}$ to $S_{t+1}^{(1)}\cdots S_{t+1}^{(C)}$ is $\prod_c\sum_{c'}H_{S_t^{(c')},S_{t+1}^{(c)}}^{(c',c)}$, following the conditional marginal probability distributions $\mathbb{P}\left(S_{t+1}^{(c)}|S_t^{(1)}\cdots S_t^{(C)}\right)$ and the independence of the conditional marginal probability distributions at time $t+1$. When the rates of state change is sufficiently small (i.e., when $\min_{S_{t+1}^{(c)},S_t^{(1)},\cdots,S_t^{(C)}}\mathbb{P}(S_{t+1}^{(c)}|S_t^{(1)}\cdots S_t^{(C)})$ is sufficiently small), approximately at most one node changes state at each time step.

We can also consider the evolution of the nodes' states in a continuous-time framework. Let $H_{i,j}^{(c',c)}$ be the rate of the event that node $c'$ in state $S_0^{(c')} = i$ causes node $c$ to jump to a different state $j \neq S_0^{(c)}$. The first occurrence time of this event (taking the current time as 0) is hence an exponential distribution:

$$\lim_{\Delta t\to 0}\underbrace{\left(1-H_{i,j}^{(c',c)}\cdot\Delta t\right)^{\lfloor t/\Delta t\rfloor}}_{\text{event not happen before } t}H_{i,j}^{(c',c)}\Delta t = H_{i,j}^{(c',c)}\exp\left(-H_{i,j}^{(c',c)}\cdot t\right)\Delta t$$

$$\Leftrightarrow \quad t \sim \text{Exponential}\left(H_{i,j}^{(c',c)}\right).$$

The first occurrence time of any such event that causes node $c$ to leave its current state is the minimum among all such the first occurrence times, and it is again an exponential distribution:

$$\mathbb{P}\left(\text{first event requesting } c \text{ to change state arrives at } t\right)$$

$$= \underset{\text{all such events}}{\prod}\mathbb{P}\left(\text{event does not happen before } t\right)$$

$$= \prod_{c'}\int_t^{\infty}H_{i,j}^{(c',c)}\exp\left(-H_{S_0^{(c')},j}^{(c',c)}\cdot\tau\right)d\tau = \prod_{c'}\exp\left(-H_{S_0^{(c')},j}^{(c',c)}\cdot t\right) = \exp\left(-\sum_{c'}H_{S_0^{(c')},j}^{(c',c)}\cdot t\right)$$

$$\Leftrightarrow \quad t \sim \text{Exponential}\left(\sum_c H_{S_0^{(c')},j}^{(c',c)}\right), \text{ where } j \neq S_0^{(c)}.$$

Given that $t$ is the first occurrence time of such events that cause node $c$ to leave its current state, the likelihood that a specific event happens is proportional to the rate of the event,

$$\mathbb{P}\left(X_1 \in (t, t+dt) \mid \min(X_1, \cdots, X_n) \in (t, t+dt)\right)$$

$$= \mathbb{P}\left(X_1 \in (t, t+dt); X_2, \cdots, X_n > t\right) \Big/ \mathbb{P}\left(\min(X_1, \cdots, X_n) \in (t, t+dt)\right)$$

$$= \lambda_1 \exp(-\lambda_1 t) \prod_{i=2}^{n} \exp(-\lambda_i t) \, dt \Big/ \left(\sum_{i=1}^{n} \lambda_i\right) \exp\left(-t \cdot \sum_{i=1}^{n} \lambda_i\right) dt$$

$$= \lambda_1 \Big/ \sum_{i=1}^{n} \lambda_i,$$

and two such occational events are unlikely to happen simultaneously.

Exponential distribution often appears in physical sciences because it is memoryless and it has the maximum entropy among all probability distributions in $[0, \infty)$ with given mean and whose density functions are positive everywhere. The exponential distribution is memoryless because it does not remember its past and behave as if the past never happened at each time: $\mathbb{P}(X > t + s \mid X > t) = \mathbb{P}(X > s)$ or

$$\mathbb{P}(X > t + s) = \mathbb{P}(X > t + s \mid X > t) \cdot \mathbb{P}(X > t) = \mathbb{P}(X > s)\mathbb{P}(X > t)$$

or specifically,     $\underbrace{\int_{t+s}^{\infty} H_{i,j}^{(c',c)} \exp\left(-H_{i,j}^{(c',c)} \cdot \tau\right) d\tau}_{=\exp\left(-H_{i,j}^{(c',c)} \cdot (t+s)\right)}$

$$= \underbrace{\int_{t}^{\infty} H_{i,j}^{(c',c)} \exp\left(-H_{i,j}^{(c',c)} \cdot \tau\right) d\tau}_{=\exp\left(-H_{i,j}^{(c',c)} \cdot t\right) \cdot} \cdot \underbrace{\int_{s}^{\infty} H_{i,j}^{(c',c)} \exp\left(-H_{i,j}^{(c',c)} \cdot \tau\right) d\tau}_{=\exp\left(-H_{i,j}^{(c',c)} \cdot s\right)}.$$

We can check that the exponential distribution has the maximum entropy with calculus of variations.

$$\mathcal{L}[p] = -\int_{\mathbb{R}^+} p(x) \log p(x) dx + \alpha \left(1 - \int_{\mathbb{R}^+} p(x) dx\right) + \beta \left(\mu - \int_{\mathbb{R}^+} x \cdot p(x) dx\right)$$

$$\frac{\partial \mathcal{L}}{\partial p} = 1 + \log p - \alpha - \beta \cdot x \overset{\text{set}}{=} 0, \quad \int_{\mathbb{R}^+} p(x) dx = 1, \quad \int_{\mathbb{R}^+} x p(x) dx = \mu$$

$$\Rightarrow \quad p(x) = \frac{1}{\mu} \exp\left(-\frac{x}{\mu}\right), \quad \exp(-1 - \alpha) = \beta, \beta = 1/\mu.$$

Asavathiratham developed the influence model to describe the dynamics of a graph that is composed of a large set of nodes (or vertices) and a set of edges that connect the nodes. In this graph, each node takes a state from its state space at any specific time, and changes its state stochastically according to both the messages sent from the other nodes through the edges and the messages sent from itself. This message-passing dynamics has much less complexity than the general dynamics in which any state combination of the nodes can lead to any other state combination. On the other hand, it provides us a quantitative language to describe our intuition on how people change

their behavior from the accumulated influences of their friends, how brainstorming hits a new idea that receives enough activation from nearby ideas, and how neurons fire based on the votes from other neurons. Asavathiratham further extended the influence model into a system where a subset of nodes and send a message to another subset of nodes and change the state combination of the latter.

In order both to compute through simulations and and to analytically express some statistics about the evolution of the states of the graph nodes, Asavathiratham specified the dynamics of the graph nodes in such a way that the marginal probability distributions of the nodes at time $t + 1$ can be estimated as a weighted sum of the marginal probability distributions at $t$. Asavathiratham found a way to translate between the weights to update the marginal probability distributions and the state transition probabilities from one state combination of all nodes to another state combination of all nodes. Hence he is able to connect the statistics of the influence process about a graph and the statistics of the Markov chain about the graph. In particular, he is able to chaim that given a Markov chain of a graph, he could create a logarithmically simpler influence process which while having different dynamics has the same asymptotic marginal probability distributions of the nodes' states.

**Definition 1** (influence model). An influence process $\left(S_t^{(c)} : c, t\right)$, where $t \subseteq \mathbb{N}$, $c = \{1, \cdots, C\}$ and $S_t^{(c)} \in \{1, \cdots, m_c\}$, identifies a family of Markov chains parameterized by row vectors $\pi_{1 \times m_c}^{(c)}$, matrix $D_{C \times C}$ and matrices $A_{m_c \times m_c}^{(c_1, c_2)}$:

$$\mathbb{P}\left(S_1^{(c)}\right) = \pi^{(c)}\left(S_1^{(c)}\right),$$

$$\mathbb{P}\left(S_{t+1}^{(c)}\right) = \sum_{c'=1}^{C} \sum_{S_t^{(c')}=1}^{m_{c'}} H_{S_t^{(c')} S_{t+1}^{(c)}}^{(c',c)} = \sum_{c'=1}^{C} D^{(c',c)} \sum_{S_t^{(c')}=1}^{m_{c'}} A_{S_t^{(c')} S_{t+1}^{(c)}}^{(c',c)}.$$

With this influence model definition, Asavathiratham simulated how a network of power plants go normal and overheated by the influences of one another in the following way:

**at time** 1 each power plant randomly chooses its initial state, $S_1^{(c)} \sim \pi^{(c)}$, and

**at** $t + 1$ each power plant randomly chooses its state indepdently according to the states of the power plants at time $t$, $\mathbb{P}\left(S_{t+1}^{(c)}\right) = \sum_{c'=1}^{C} \sum_{S_t^{(c')}=1}^{m_{c'}} H_{S_t^{(c')} S_{t+1}^{(c)}}^{(c',c)}.$

Once he specifies a way to simulate a influence model, Asavathiratham picks a specific Markov chain from the family of Markov chains that are compatible with the influence model specification. For instance, the Markov chain describing the state evolution of the power plant network has the state transition probability

$$\mathbb{P}\left(S_{t+1}^{(1)} \cdots S_{t+1}^{(C)} | S_t^{(1)} \cdots S_t^{(C)}\right) = \prod_c \mathbb{P}\left(S_{t+1}^{(c)} | S_t^{(1)} \cdots S_t^{(C)}\right) = \prod_c \sum_{c'} H_{S_t^{(c')} S_{t+1}^{(c)}}^{(c',c)}.$$

On the other hand, Asavathiratham could simulate the power plant network with other Markov chain specifications and still conformed to the influence model to be simulated. The differently

simulated versions of the influence model will have different dynamics about the whole power plant network, yet they have the same dynamics about any specific power plant. A simple way to simulate with different dynamics about the power plant network is to correlate the conditional probability distributions $\mathbb{P}\left(S_{t+1}^{(c)}|S_t^{(1)}\cdots S_t^{(C)}\right)$.

Asavathiratham related the Markov chain description of graph dynamics and the influence model description with an **event matrix**. Let us denote the state transition matrix (or state transition kernel) of the graph as $\left(\mathbb{P}\left(S_t^{(1)}\cdots S_t^{(C)} \to S_{t+1}^{(1)}\cdots S_{t+1}^{(C)}\right)\right)_{\prod m_c \times \prod m_c}$. This matrix has $\prod m_c$ rows and $\prod m_c$ columns, and each row/column is indexed by a s state combination of the graph nodes. Let us write the **influence matrix** corresponding to the influence model as $\left(H_{S_t^{(c')}S_t^{(c)}}^{(c',c)}\right)_{\sum m_c \times \sum m_c}$. This matrix has $\sum m_c$ rows and $\sum m_c$ columns, and each row/column is indexed by a state of a node. The event matrix $\left(B\left(S_t^{(1)}\cdots S_t^{(C)},(c,i)\right) \overset{\text{def.}}{=} \left(S^{(c)} \overset{?}{=} i\right)\right)_{\prod m_c \times \sum m_c}$ has $\prod m_c$ rows and $\sum m_c$ columns. In this event matrix, the row indexed by $S_t^{(1)}\cdots S_t^{(C)}$ and the column indexed by $(c,i)$ is one if $S_t^{(c)} = i$ and zero otherwise. Asavathiratham showed that

$$\sum_{S_t=\cdots i\cdots, S_{t+1}=\cdots j\cdots} B(S_t,(c',i)) \cdot \mathbb{P}(S_t \to S_{t+1}) \cdot B(S_{t+1},(c,j)) = H_{i,j}^{(c',c)}.$$

Asavathiratham's influence modeling provides us a good example of studying the stochastic process of a node by relating the stochastic process of the whole graph to a Markov chain. His approach also tells us through an eigen-decomposition argument that we can seek to get the statistics of a computationally intractable network by working with a much simpler network, which has the same statistics that we are interested in and very different dynamics otherwise. For instance, computing the marginal probability distirbution of the asymptotic state of a power plant in a power plant network could be intractable due to the complexity of the network dynamics, and Asavathiratham showed us that he can compute the same distribution by working with a much simpler network in which the power plants independently change their states through the influences of one another. We make many approximations too in studying knowledge networks and social networks due to the sparsity of the available human behavior data and the diversity and randomness of human behavior. We learned through Asavathiratham's story that we can sometimes get a good estimation of our interested statistics through a good selection of an approximation network. On the other hand, we should noticed that Asavathiratham's influence process represents a family of Markov chains with different dynamics and probability measures, hence it is an ill-posed problem to talk about the (log) likelihood of the Asavathiratham's influence process.

The influence model defined by Asavathiratham is not the only way to describe a large system with interacting parts. The model has a limitation about "influence" since different states of a chain are required to have the same influence to another chain, and we can define another influence model without this limitation.

## 2.2 Completely Factorized Variational Inference

An influence model is a Markov chain $\left(S_t^{(c)} \in \{1, \cdots, m_c\}, U_t^{(c)} \in \{1, \cdots, C\}\right)_{t \subseteq \mathbb{N}}^{c = \{1, \cdots, C\}}$ param-
eterized by row vectors $\pi_{1 \times m_c}^{(c)}$, matrix $D_{C \times C}$ and matrices $A_{m_c \times m_c}^{(c_1, c_2)}$.

$$\mathbb{P}\left(\left\{S_t^{(c)}, U_t^{(c)} : c = \{1, \cdots, C\}, t = \{1, \cdots, T\}\right\}\right) = \prod_{c=1}^{C} \pi_{S_1^{(c)}}^{(c)} \cdot \prod_{t=1}^{T-1} \prod_{c=1}^{C} D_{U_t^{(c)}, c} A_{S_t^{\left(U_t^{(c)}\right)}, S_{t+1}^{(c)}}^{\left(U_t^{(c)}, c\right)} \cdot$$

with latent state

$$\mathbb{P}\left(\left\{S_t^{(c)}, U_t^{(c)}, Y_t^{(c)} : c \in \{1, \cdots, C\}, t \in \{1, \cdots, T\}\right\}\right) \tag{2.1}$$

$$= \prod_{c=1}^{C} \pi_{S_1^{(c)}}^{(c)} \cdot \prod_{t=1}^{T-1} \prod_{c=1}^{C} H_{S_t^{(U_t(c))}, S_{t+1}^{(c)}}^{(U_t(c), c)} \cdot \prod_{t=1}^{T} \prod_{c=1}^{C} \mathbb{P}\left(Y_t^{(c)} | S_t^{(c)}\right)$$

$$= \prod_{c=1}^{C} \prod_{i=1}^{m_c} \left(\pi_i^{(c)}\right)^{S_{1,i}^{(c)}} \cdot \prod_{t, c_1, c, i, j = 1}^{T-1, C, C, m_{c_1}, m_c} \left(H_{i,j}^{(c_1, c)}\right)^{U_t(c_1, c) S_{t,i}^{(c_1)} S_{t+1,j}^{(c)}} \cdot \prod_{t, c, i = 1}^{T, C, m_c} \mathbb{P}\left(Y_t^{(c)} | S_t^{(c)} = i\right).$$

Hence the log probability is

$$\log \mathbb{P}\left(\left\{S_t^{(c)}, U_t^{(c)}, Y_t^{(c)} : c \in \{1, \cdots, C\}, t \in \{1, \cdots, T\}\right\}\right) \tag{2.2}$$

$$= \sum_{c=1}^{C} \sum_{i=1}^{m_c} S_{1,i}^{(c)} \log \pi_i^{(c)} + \sum_{t, c, c_1, i, j} U_t(c_1, c) S_{t,j}^{(c_1)} S_{t+1,i}^{(c)} \log H_{i,j}^{(c_1, c)} + \sum_{t, c, i} S_{t,i}^{(c)} \log \mathbb{P}\left(Y_t^{(c)} | S_t^{(c)} = i\right)$$

$$= \sum_{c, i = 1}^{C, m_c} S_{1,i}^{(c)} \log \pi_i^{(c)} + \sum_{t, c, i = 1}^{T, C, m_c} S_{t,i}^{(c)} \log \mathbb{P}\left(Y_t^{(c)} | S_t^{(c)} = i\right) +$$

$$+ \sum_{t, c, c_1, i, j = 1}^{T-1, C, C, m_{c_1}, m_c} U_t(c_1, c) S_{t,i}^{(c_1)} S_{t+1,j}^{(c)} \log A_{i,j}^{(c_1, c)} + \sum_{t, c, c_1 = 1}^{T-1, C, C} U_t(c_1, c) \log D^{(c_1, c)}.$$

The probability and the log probability in the mean-field approximation are respectively

$$\mathbb{Q}\left(\left\{S_t^{(c)}, U_t^{(c)}, Y_t^{(c)} : c \in \{1, \cdots, C\}, t \in \{1, \cdots, T\}\right\}\right) = \prod_{t, c = 1}^{T, C} \mathbb{Q}(S_t^{(c)}) \cdot \prod_{t, c = 1}^{T, C} \mathbb{Q}(U_t^{(c)}) =$$

$$= \prod_{t=1}^{T} \prod_{c=1}^{C} \prod_{i=1}^{m_c} \left(\sigma_{t,i}^{(c)}\right)^{S_{t,i}^{(c)}} \cdot \prod_{t=1}^{T} \prod_{c=1}^{C} \prod_{c_1=1}^{C} \left(\iota_t(c_1, c)\right)^{U_t(c_1, c)},$$

$$\log \mathbb{Q}\left(\left\{S_t^{(c)}, U_t^{(c)}, Y_t^{(c)} : c \in \{1, \cdots, C\}, t \in \{1, \cdots, T\}\right\}\right)$$

$$= \sum_{t=1}^{T} \sum_{c=1}^{C} \sum_{i=1}^{m_c} S_{t,i}^{(c)} \log \sigma_{t,i}^{(c)} + \sum_{t=1}^{T} \sum_{c=1}^{C} \sum_{c_1=1}^{C} U_t(c_1, c) \log \iota_t(c_1, c)$$

Hence the KL divergence

$$KL(\mathbb{Q}\|\mathbb{P}) = \mathbb{E}_\mathbb{Q}(\log\mathbb{Q}) - \mathbb{E}_\mathbb{Q}(\log\mathbb{P})$$

$$= \sum_{t=1}^{T}\sum_{c=1}^{C}\sum_{i=1}^{m_c} \sigma_{t,i}^{(c)} \log\sigma_{t,i}^{(c)} + \sum_{t=1}^{T}\sum_{c=1}^{C}\sum_{c_1=1}^{C} \iota_t(c_1,c)\log\iota_t(c_1,c)$$

$$- \sum_{c,i=1}^{C,m_c} \sigma_{1,i}^{(c)}\log\pi_i^{(c)} - \sum_{t,c,c_1,i,j=1}^{T-1,C,C,m_c,m_{c_1}} \iota_t(c_1,c)\sigma_{t,i}^{(c_1)}\sigma_{t+1,j}^{(c)}\log H_{i,j}^{(c_1,c)} - \sum_{t,c,i,j=1}^{T,C,m_c,n_c} \sigma_{t,i}^{(c)}Y_{t,j}^{(c)}\log B_{i,j}.$$

It is a convex optimization problem with constraints to find the best $\mathbb{Q}$ that minimizes $KL(\mathbb{Q}\|\mathbb{P})$. In particular, the parameters for $\mathbb{Q}$ can be solve by fixed point iteration that is in essence an EM algorithm: Taking partial derivatives.

$$\frac{\partial KL}{\partial\sigma_{1,i}^{(c)}} = 1 + \log\sigma_{1,i}^{(c)} - \log\pi_{1,i}^{(c)} - \sum_{c_2=1}^{C}\sum_{j=1}^{m_{c_2}} \iota_1(c,c_2)\sigma_{2,j}^{(c_2)}\log H_{i,j}^{(c,c_2)} - \log\mathbb{P}(Y_1^{(c)}|S_1^{(c)}=i)$$

$$\frac{\partial KL}{\partial\sigma_{t>1,i}^{(c)}} = 1 + \log\sigma_{t,i}^{(c)} + \sum_{c_1,k=1}^{C,m_{c_1}} \iota_{t-1}(c_1,c)\sigma_{t-1,k}^{(c_1)}\log H_{k,i}^{(c_1,c)} + \sum_{c_2,j=1}^{C,m_{c_2}} \iota_t(c,c_2)\sigma_{t+1,j}^{(c_2)}\log H_{i,j}^{(c,c_2)} + \log\mathbb{P}(Y_t^{(c)}|S_t^{(c)}=i)$$

$$\frac{\partial KL}{\partial\iota_t(c_1,c)} = 1 + \log\iota_t(c_1,c) + \sum_{,i,j=1}^{m_c,m_{c_1}} \sigma_{t,i}^{(c_1)}\sigma_{t+1,j}^{(c)}\log H_{i,j}^{(c_1,c)}$$

Hence

$$\sigma_{1,i}^{(c)}\text{ (new)} = \frac{\pi_{1,i}^{(c)}\cdot\prod_{c_2,j}\left(H_{i,j}^{(c,c_2)}\right)^{\iota_1(c,c_2)\sigma_{2,j}^{(c_2)}}\mathbb{P}\left(Y_1^{(c)}|S_1^{(c)}=i\right)}{\sum_i\pi_{1,i}^{(c)}\cdot\prod_{c_2,j}\left(H_{i,j}^{(c,c_2)}\right)^{\iota_1(c,c_2)\sigma_{2,j}^{(c_2)}}\mathbb{P}\left(Y_1^{(c)}|S_1^{(c)}=i\right)}$$

$$\sigma_{t>1,i}^{(c)}\text{ (new)} = \frac{\prod_{c_1,k=1}^{C,m_{c_1}}\left(H_{k,i}^{(c_1,c)}\right)^{\iota_{t-1}(c_1,c)\sigma_{t-1,k}^{(c_1)}}\prod_{c_2,j=1}^{C,m_{c_2}}\left(H_{i,j}^{(c,c_2)}\right)^{\iota_t(c,c_2)\sigma_{t+1,j}^{(c_2)}}\mathbb{P}\left(Y_t^{(c)}|S_t^{(c)}=i\right)}{\sum_i\prod_{c_1,k=1}^{C,m_{c_1}}\left(H_{k,i}^{(c_1,c)}\right)^{\iota_{t-1}(c_1,c)\sigma_{t-1,k}^{(c_1)}}\prod_{c_2,j=1}^{C,m_{c_2}}\left(H_{i,j}^{(c,c_2)}\right)^{\iota_t(c,c_2)\sigma_{t+1,j}^{(c_2)}}\mathbb{P}\left(Y_t^{(c)}|S_t^{(c)}=i\right)}$$

$$\iota_t(c_1,c)\text{ (new)} = \prod_{i,j=1}^{m_c,m_{c_1}}\left(H_{i,j}^{(c,c_2)}\right)^{\sigma_{t,i}^{(c_1)}\sigma_{t+1,j}^{(c)}} \bigg/ \sum_{c_1=1}^{C}\prod_{,i,j=1}^{m_c,m_{c_1}}\left(H_{i,j}^{(c,c_2)}\right)^{\sigma_{t,i}^{(c_1)}\sigma_{t+1,j}^{(c)}}$$

The parameters $\{\sigma_{t,i}^{(c)} : c,t,i\}$ and $\{\iota_t(c_1,c) : t,c_1,c\}$ of the mean-field approximation $\mathbb{Q}$ adapt to the posterior probability measure $\mathbb{P}(\bullet|\{y_t^{(c)} : c,t\})$ by balancing between the incompatible needs of minimizing $\mathbb{E}_\mathbb{Q}(\log\mathbb{Q})$ and maximizing $\mathbb{E}_\mathbb{Q}(\log\mathbb{P})$ in a procedure to minimize the KL divergence $KL(\mathbb{Q}\|\mathbb{P})$. They minimize $\mathbb{E}_\mathbb{Q}(\log\mathbb{Q})$ with $\sigma_{t,i}^{(c)} = 1/m_c$ and $\iota_t(c_1,c) = 1/C$, and corresponding to this parameterization every latent state path $\{S_t^{(c)}, U_t(c) : c,t\}$ is equally likely. They maximize $\mathbb{E}_\mathbb{Q}(\log\mathbb{P})$ by assigning all probability weight to one specific latent state path $\{S_t^{(c)}, U_t(c) : c,t\}$. The parameters $\sigma_{t,\bullet}^{(c)}$ and $\iota_t(\bullet,c)$ in a minimization procedure need to adapt to the observation $Y_t^{(c)}$, need to match $\sigma_{t\pm1,\bullet}^{(c_1)}$ and $\iota_{t\pm1}(\bullet,c_1)$ through the influence parameters $\left(H_{i,j}^{(c_1,c)}\right)$, and need to match $\sigma_{t\pm\Delta t,\bullet}^{(c_2)}$ and $\iota_{t\pm\Delta t}(\bullet,c_2)$ for $\Delta t > 1$ through $\sigma_{t\pm1,\bullet}^{(c_1)}$ and $\iota_{t\pm1}(\bullet,c_1)$ and the influence parameters.

## 2.3 Structured Variation Inference

Let us use the "structure" that I have already had to approximate the probability measure of the latent state influence model.

I developed an approximation algorithm for the E-step of the latent state influence model. This algorithm works sufficiently well for the applications that I have investigated. It is based on the following relationships among the marginal probability statistics of the latent states.

Let the marginal forward statistics $\alpha_t^{(c)}(s_t^{(c)})$, the marginal backward statistics $\beta_t^{(c)}(s_t^{(c)})$, the marginal one-slice statistics $\gamma_t^{(c)}(s_t^{(c)})$, the marginal two-slice statistics $\xi_{t \to t+1}^{(c_1,c_2)}(s_t^{(c_1)}, s_{t+1}^{(c_2)})$ of a latent structure influence model be

$$\alpha_t^{(c)}(S_t^{(c)}) = \mathbb{P}\left(S_t^{(c)}, \left\{Y_{t_1}^{(c_1)} : c_1 = 1, \cdots, C; t_1 = 1, \cdots, t\right\}\right)$$

$$\beta_t^{(c)}(S_t^{(c)}) = \mathbb{P}\left(\left\{Y_{t_1}^{(c_1)} : c_1 = 1, \cdots, C; t_1 = t+1, \cdots, T\right\} | S_t^{(c)}, \left\{\alpha_t^{(c_1)} : c_1 \neq c\right\}\right)$$

$$\gamma_t^{(c)}(S_t^{(c)}) = \mathbb{P}\left(S_t^{(c)} | \left\{Y_{t_1}^{(c_1)} : c_1 = 1, \cdots, C; t_1 = 1, \cdots, T\right\}\right)$$

$$\xi_{t \to t+1}^{(c_1,c_2)}(S_t^{(c_1)}, S_{t+1}^{(c_2)}) = \mathbb{P}\left(S_t^{(c_1)} S_{t+1}^{(c_2)} | \left\{Y_{t_1}^{(c_1)} : c_1 = 1, \cdots, C; t_1 = 1, \cdots, T\right\}\right).$$

We can deduce the following inductive relationship among the forward statistics from the structure of the latent state influence model:

$$\alpha_1^{(c)}(S_1^{(c)}) = \overbrace{\sum_{\left\{S_1^{(c')}:c';\text{fix } S_1^{(c)}\right\}} \mathbb{P}\left(\{S_1^{(c')}, Y_1^{(c')} : c'\}\right)}^{\text{by definition of } \alpha_1^{(c)}} = \overbrace{\sum_{\left\{S_1^{(c')}:c';\text{fix } S_1^{(c)}\right\}} \prod_{c'} \pi_{S_1^{(c')}}^{(c')} \mathbb{P}\left(Y_1^{(c')} | S_1^{(c')}\right)}^{\text{by influence model structure}} =$$

$$= \overbrace{\prod_{c' \neq c} \sum_{S_1^{(c')}} \pi_{S_1^{(c')}}^{(c')} \mathbb{P}\left(Y_1^{(c')} | S_1^{(c')}\right) \cdot \pi_{S_1^{(c)}}^{(c)} \mathbb{P}\left(Y_1^{(c)} | S_1^{(c)}\right)}^{\text{factorization}}$$

$$\alpha_{t+1}^{(c)}(S_{t+1}^{(c)}) = \overbrace{\mathbb{P}\left(\left\{Y_{t+1}^{(c')} : c'\right\} | S_{t+1}^{(c)}\right) \mathbb{P}\left(S_t^{(c)}, \left\{Y_\tau^{(c')} : c', \tau = 1, \cdots, t\right\}\right)}^{\text{by definition of } \alpha_t^{(c)}}$$

$$= \mathbb{P}\left(\left\{Y_{t+1}^{(c')} : c'\right\} | S_{t+1}^{(c)}\right) \cdot \underbrace{\sum_{c_1} \sum_{S_t^{(c_1)}} H_{S_t^{(c_1)} S_{t+1}^{(c)}}^{(c_1,c)} \overbrace{\frac{\mathbb{P}\left(S_{t-1}^{(c_1)}, \{Y_\tau^{(c')} : c', \tau \leq t\}\right)}{\mathbb{P}\left(\{Y_\tau^{(c')} : c', \tau \leq t\}\right)}}^{\text{influence model structure}} \mathbb{P}\left(\{Y_\tau^{(c')} : c', \tau \leq t\}\right)}_{\alpha_t^{(c)}(S_t^{(c)})}.$$

Known the posterior marginal distributions $\alpha_t^{(c)}$, we can proceed to deduce the posterior backward parameters by the same reasoning in the deduction of the forward statistics, and the other statistics:

$$\beta_t^{(c)}\left(S_t^{(c)}\right) = \sum_{S_T^{(c_1)}} \mathbb{P}\left(S_T^{(c_1)}, \left\{Y_\tau^{(c')} : c'; t < \tau \leq T\right\} | S_t^{(c)}, \left\{\alpha_t^{(c')} : c' \neq c\right\}\right)$$

$$\gamma_t^{(c)}\left(S_t^{(c)}\right) = \alpha_t^{(c)}\left(S_t^{(c)}\right) \cdot \beta_t^{(c)}\left(S_t^{(c)}\right)$$

$$\xi_{t \to t+1}^{(c_1,c_2)}(S_t^{(c_1)}, S_{t+1}^{(c_2)}) = \alpha_t^{(c_1)}\left(S_t^{(c_1)}\right) H_{S_t^{(c_1)} S_{t+1}^{(c_2)}}^{(c_1,c_2)} \mathbb{P}\left(\left\{Y_{t+1}^{(c')} : c'\right\} | S_{t+1}^{(c)}\right) \cdot \beta_{t+1}^{(c_2)}\left(S_{t+1}^{(c_2)}\right).$$

The forward, backward, one-slice and two-slice statistics seem to give an exact inference algorithm for the latent state influence model. But when we look behind the nice forms of the statistics, we find that there is generally not an exact way of computing $\mathbb{P}\left(\left\{Y_{t+1}^{(c')} : c'\right\} | S_{t+1}^{(c)}\right)$ without coping with the exploding number of terms involving $\left(S_{t+1}^{(1)} \cdots S_{t+1}^{(C)}\right)$. In the approximation algorithm that I developed, I assumed that

$$\alpha_{t+1}^{(c)}(S_{t+1}^{(c)}) \overset{\text{approx.}}{\propto} \mathbb{P}\left(Y_{t+1}^{(c)} | S_{t+1}^{(c)}\right) \cdot \sum_{c_1} \sum_{S_t^{(c_1)}} H_{S_t^{(c_1)} S_{t+1}^{(c)}}^{(c_1,c)} \alpha_t^{(c_1)}\left(S_t^{(c_1)}\right).$$

We correctly computed

$$\mathbb{P}\left(S_{t+1}^{(c)} | \left\{Y_\tau^{(c)} : c, \tau \leq t\right\}\right) \quad = \sum_{\left\{S_{t+1}^{(c')} : c'; \text{fix } S_{t+1}^{(c)}\right\}} \mathbb{P}\left(\left\{S_{t+1}^{(c')} : c'\right\} | \left\{Y_\tau^{(c')} : c', \tau \leq t\right\}\right)$$

through the definition of the influence model. On the other hand, the latent states $\left\{S_{t+1}^{(c)} : c\right\}$ are correlated, and hence both $Y_{t+1}^{(c')}$ with $c \neq c'$ and $Y_{t+1}^{(c)}$ contain information for updating $S_{t+1}^{(c)}$, and

$$\mathbb{P}\left(S_{t+1}^{(c)} | \left\{Y_\tau^{(c')} : c', \tau \leq t\right\}\right) \cdot \mathbb{P}\left(Y_{t+1}^{(c)} | S_{t+1}^{(c)}\right)$$
$$\neq \sum_{\left\{S_{t+1}^{(c')} Y_{t+1}^{(c')} : c'; \text{ fix } S_{t+1}^{(c)}, Y_{t+1}^{(c)}\right\}} \mathbb{P}\left(\left\{S_{t+1}^{(c')} : c'\right\} | \left\{Y_\tau^{(c')} : c', \tau \leq t\right\}\right) \cdot \prod_{c'} \mathbb{P}\left(Y_{t+1}^{(c')} | S_{t+1}^{(c')}\right).$$

Below I give the variational method that adapt $h_{t,i}^{(c)}$ to $\mathbb{P}\left(\left\{Y_t^{(c')} : c'\right\} | S_t^{(c)} = i\right)$. Recall that the likelihood function for a latent state influence model with known latent state sequences and unknown "influencers" is

$$\mathbb{P}\left(\left\{S_t^{(c)}, Y_t^{(c)} : c, t\right\}\right) = \prod_{c=1}^C \mathbb{P}\left(Y_1^{(c)} | S_1^{(c)}\right) \pi_{S_1^{(c)}}^{(c)} \cdot \prod_{t=1}^{T-1} \prod_{c=1}^C \mathbb{P}\left(Y_{t+1}^{(c)} | S_{t+1}^{(c)}\right) \sum_{c_1=1}^C H_{S_t^{(c_1)}, S_{t+1}^{(c)}}^{(c_1,c)} =$$

$$= \prod_{c=1}^C \prod_{i=1}^{m_c} \left(\mathbb{P}\left(Y_1^{(c)} | S_1^{(c)} = i\right)\right)^{S_{1,i}^{(c)}} \left(\pi_i^{(c)}\right)^{S_{1,i}^{(c)}} \cdot \prod_{t=1}^{T-1} \prod_{c=1}^C \prod_{i=1}^{m_c} \left(\mathbb{P}\left(Y_{t+1}^{(c)} | S_{t+1}^{(c)} = i\right)\right)^{S_{t+1,i}^{(c)}} \prod_{S_t} \left(\sum_{c_1=1}^C H_{S_t^{(c_1)}, S_{t+1}^{(c)}}^{(c_1,c)}\right)^{S_t}$$

$$\log \mathbb{P}\left(\left\{S_t^{(c)}, Y_t^{(c)} : c, t\right\}\right) = \sum_{c=1}^C \sum_{i=1}^{m_c} S_{1,i}^{(c)} \log \mathbb{P}\left(Y_1^{(c)} | S_1^{(c)} = i\right) + \sum_{c=1}^C \sum_{i=1}^{m_c} S_{1,i}^{(c)} \log \pi_i^{(c)} +$$

$$\sum_{t,c,i=1}^{T-1,C,m_c} S_{t+1,i}^{(c)} \log \mathbb{P}\left(Y_{t+1}^{(c)} | S_{t+1}^{(c)} = i\right) + \sum_{t,c,i=1}^{T-1,C,m_c} \sum_{S_t,\cdots} S_{t,\cdots} \log \sum_{c_1=1}^C H_{S_t^{(c_1)}, S_{t+1}^{(c)}}^{(c_1,c)} \cdot$$

The structural approximation that I just described translates into the following likelihood function.

$$\mathbb{Q}\left(\left\{S_t^{(c)} : c, t\right\}\right) = \prod_{c=1}^C \frac{\mathbb{P}\left(Y_1^{(c)} | S_1^{(c)}\right) \pi_{S_1^{(c)}}^{(c)}}{\sum_{S_1^{(c)}} \mathbb{P}\left(Y_1^{(c)} | S_1^{(c)}\right) \pi_{S_1^{(c)}}^{(c)}} \cdot \prod_{t=1}^{T-1} \prod_{c=1}^C \mathbb{Q}\left(S_t^{(c)} = i\right)$$

$$\mathbb{Q}\left(S_t^{(c)} = i\right) = \tilde{\alpha}_t^{(c)}(i) = \frac{\alpha_t^{(c)}(i)}{\sum_j \alpha_t^{(c)}(j)} = \frac{h_{t,i}^{(c)} \cdot \sum_{c_1} \sum_k \tilde{\alpha}_{t-1}^{(c_1)}(k) H_{k,i}^{(c_1,c)}}{\sum_j h_{t,j}^{(c)} \cdot \sum_{c_1} \sum_k \tilde{\alpha}_{t-1}^{(c_1)}(k) H_{k,j}^{(c_1,c)}}$$

To understand the probability measure $\mathbb{Q}$ define by the structual approximation, we treat $\mathbb{Q}$ as a product of many factors involving $\{S_t^{(c)}, S_{t+1}^{(c)} : c\}$ that compute the prior probability distribution of $\left\{S_{t+1}^{(c)} : c\right\}$ from $\left\{S_{t+1}^{(c)} : c\right\}$ and that incorporate the evidence in the observations $\left\{Y_{t+1}^{(c)} : c\right\}$. These factors are different for $\mathbb{Q}$ and for $\mathbb{P}$: The part that computes the prior probability distribution is the same, but the part that incorporate evidence is different. Since $\sum_S \mathbb{P}\left(\left\{S_t^{(c)}, Y_t^{(c)} : c, t\right\}\right) = \mathbb{P}\left(\left\{Y_t^{(c)} : c, t\right\}\right)$, this probability measure is compatible with the approximation that we used.

The KL-divergence can be computed as

$$\log \mathbb{Q}\left(\left\{S_t^{(c)}, : c, t\right\}\right) - \log \mathbb{P}\left(\left\{S_t^{(c)}, Y_t^{(c)} : c, t\right\}\right)$$

$$= \sum_{t=1}^{T-1} \sum_{c=1}^{C} \left(\sum_{i=1}^{m_c} S_{t+1,i}^{(c)} \left(\log h_{t+1,i}^{(c)} - \log \mathbb{P}\left(Y_{t+1}^{(c)}|S_{t+1}^{(c)} = i\right)\right) - \underbrace{\log \sum_j h_{t+1,j}^{(c)} \sum_{c_1} \sum_i \tilde{\alpha}_t^{(c_1)}(i) H_{i,j}^{(c_1,c)}}_{\text{unrelated to } h_{t,i}^{(c)}}\right),$$

$$KL(\mathbb{Q}\|\mathbb{P}) = \mathbb{E}_{\mathbb{Q}} \log \mathbb{Q}\left(\left\{S_t^{(c)}, : c, t\right\}\right) - \log \mathbb{P}\left(\left\{S_t^{(c)}, Y_t^{(c)} : c, t\right\}\right) + \log\left(\left\{Y_t^{(c)} : c, t\right\}\right)$$

$$= \sum_{t=1}^{T-1} \sum_{c=1}^{C} \left(\sum_{i=1}^{m_c} \tilde{\alpha}_{t+1}^{(c)}(i) \cdot \left(\log h_{t+1,i}^{(c)} - \log \mathbb{P}\left(Y_{t+1}^{(c)}|S_{t+1}^{(c)} = i\right)\right) - \log \sum_j h_{t+1,j}^{(c)} \sum_{c_1} \sum_i \tilde{\alpha}_t^{(c_1)}(i) H_{i,j}^{(c_1,c)}\right).$$

Setting the derivatives of $KL(\mathbb{Q}\|\mathbb{P})$ over $h_{t,i}^{(c)}$ to be 0, we get

$$\frac{\partial KL}{\partial h_{t,i}^{(c)}} = \sum_{k=1}^{m_c} \frac{\partial \tilde{\alpha}_t^{(c)}(k)}{\partial h_{t,i}^{(c)}} \left(\log h_{t,k}^{(c)} - \log \mathbb{P}\left(Y_t^{(c)}|S_t^{(c)} = k\right)\right) + \frac{\tilde{\alpha}_{t,i}^{(c)}}{h_{t,i}^{(c)}} - \underbrace{\frac{\sum_{c_1} \sum_k \tilde{\alpha}_{t-1}^{(c_1)}(k) H_{k,i}^{(c_1,c)}}{\sum_j h_{t,j}^{(c)} \sum_{c_1} \sum_k \tilde{\alpha}_{t-1}^{(c_1)}(k) H_{k,j}^{(c_1,c)}}}_{\tilde{\alpha}_{t,i}^{(c)}/h_{t,i}^{(c)}}$$

$$\Rightarrow \quad h_{t,k}^{(c)} = \mathbb{P}\left(Y_t^{(c)}|S_t^{(c)} = k\right), \forall c, t, k.$$

We know that a minimum point $\left\{h_{t,k}^{(c)} : c, t, k\right\}$ exists because $KL(\mathbb{Q}\|\mathbb{P})$ is lower bounded by 0.

We can define a probability measure with $\mathbb{Q}\left(S_t^{(c)} = i\right) = \gamma_t^{(c)}(i)$ from our approximation, and get better estimation of $\mathbb{P}\left(S_t^{(c)} = i \mid \left\{Y_t^{(c)} : c, t\right\}\right)$ and other statistics. This improvement involves more complicated update of $h_{t,k}^{(c)}$ and I will no pursue this path further. There are other possible structural approximations to the latent state influence model.

Sometimes we want to know who-influence-who from the state sequences of the different interacting processes in the influence model and the influence matrix. I give an estimation algorithm for finding the "influencers" below in the case that the "influencers" $U$ are independent of each other and are independent of the influence model states $S$. This algorithm is again based on minimizing

the K-L divergence. The approximate distribution and its logarithm are respectively

$$\mathbb{Q}(\{S_t^{(c)}, U_t^{(c)} : c, t\}) = \mathbb{Q}(\{U_t^{(c)} : c, t\}) \cdot \mathbb{Q}(\{S_t^{(c)} : c, t\}) = \mathbb{Q}(\{S_t^{(c)} : c, t\}) \cdot \prod_t \prod_c \mathbb{Q}(U_t^{(c)})$$

where $\mathbb{Q}\left(U_t^{(c)}\right) = \iota_t\left(U_t^{(c)}, c\right) = \prod_{c_1=1}^{C} (\iota_t(c_1, c))^{U_t(c_1, c)}$

$$\log \mathbb{Q}(\{S_t^{(c)}, U_t^{(c)} : c, t\}) = \sum_{t=1}^{T} \sum_{c=1}^{C} \sum_{c_1=1}^{C} U_t(c_1, c) \cdot \log \iota_t(c_1, c) + \log \mathbb{Q}\left(\left\{S_t^{(c)} : c, t\right\}\right).$$

Setting to 0 the partial derivatives over $\iota_t(c_1, c)$ of the Lagrangian involving the KL divergence and the constraints $\sum_{c_1} \iota_t(c_1, c) = 1$

$$KL(\mathbb{Q}\|\mathbb{P}) = \mathbb{E}_{\mathbb{Q}} \log \mathbb{Q} - \mathbb{E}_{\mathbb{Q}} \log \mathbb{P}$$

$$= \sum_{t,c,c_1} \underbrace{\iota_t(c_1, c)}_{\mathbb{E}_{\mathbb{Q}}(U(c_1,c))} \cdot \log \iota_t(c_1, c) - \sum_{t,c_1,c,i,j} \iota_t(c_1, c) \mathbb{E}_{\mathbb{Q}}\left(S_{t,i}^{(c_1)} S_{t+1,j}^{(c)}\right) \cdot \log H_{i,j}^{(c_1,c)} + \text{irrelavant terms}$$

$$\frac{\partial \left(KL - \sum_{t,c} \omega^{(c)} \left(1 - \sum_{c_1} \iota_t(c_1, c)\right)\right)}{\partial \iota_t(c_1, c)} = 1 + \log \iota_t(c_1, c) - \mathbb{E}_{\mathbb{Q}}\left(S_{t,i}^{(c_1)} S_{t+1,j}^{(c)}\right) \cdot \log H_{i,j}^{(c_1,c)}$$

we get the following softmax iteration to calculate $\iota_t$.

$$\iota_t(c_1, c)^{(\text{new})} = \prod_{i,j=1}^{m_c, m_{c_1}} \left(H_{i,j}^{(c,c_2)}\right)^{\sigma_{t,i}^{(c_1)} \sigma_{t+1,j}^{(c)}} \bigg/ \sum_{c_1=1}^{C} \prod_{,i,j=1}^{m_c, m_{c_1}} \left(H_{i,j}^{(c,c_2)}\right)^{\mathbb{E}_{\mathbb{Q}}\left(S_{t,i}^{(c_1)} S_{t+1,j}^{(c)}\right)}$$

## 2.4  Parameter estimation

Review parameter estimation for HMM.

Let us recall how we use the maximum (log) likelihood method to estimate with state transition matrix $(p_{i \to j})_{i,j}$ and the initial state distribution $(\pi_i)_i$ of a hidden Markov model (HMM) with observation distribution $\mathbb{P}(Y_t|S_t)$. The likelihood and log-likelihood of a Markov chain of $(S_t)_t$, are respectively

$$\mathbb{P}(\{S_t, Y_t\}) = \mathbb{P}(S_1) \prod_t \mathbb{P}(S_{t+1}|S_t) \prod_t \mathbb{P}(Y_t|S_t) = \pi_{S_1} \prod_t p_{S_t \to S_{t+1}} \prod_t \mathbb{P}(Y_t|S_t) =$$

$$= \prod_i (\pi_i)^{S_{1,i}} \prod_t \prod_i \prod_j (p_{i \to j})^{S_{t,i} S_{t+1,j}} \prod_t (\mathbb{P}(Y_t|S_t = i))^{S_{t,i}}, \text{ and}$$

$$\log \mathbb{P}(\{S_t, Y_t\}) = \sum_i S_{1,i} \cdot \log \pi_i + \sum_{t,i,j} S_{t,i} S_{t+1,j} \cdot \log p_{i \to j} + \sum_{t,i} S_{t,i} \log \mathbb{P}(Y_t|S_t = i).$$

The constraints to maximizing the log likelihood are $\sum_j p_{i \to j} = 1$ for all $i$ and $\sum_i \pi_i = 1$. The Lagrangian is hence $\mathcal{L} = \log \mathbb{P} + \sum_i \mu_i \left(1 - \sum_j p_{i \to j}\right) + \nu \left(1 - \sum_i \pi_i\right)$. We can get the maximum

likelihood estimation of the parameters of the Markov chain by setting the partial derivatives of the Lagrangian over the parameters to 0 and checking that the second order partial derivatives of the loglikelihood over the parameters are negative.

$$\frac{\partial \mathcal{L}}{\partial p_{i \to j}} = \frac{\sum_t S_{t,i} S_{t+1,j}}{p_{i \to j}} - \mu_i \overset{\text{set}}{=} 0 \forall j, \sum_j p_{i \to j} = 1 \quad \Rightarrow \quad \mu_i = \sum_t S_{t,i}, p_{i \to j} = \frac{\sum_t S_{t,i} S_{t+1,j}}{\sum_t S_{t,i}}$$

$$\frac{\partial \mathcal{L}}{\partial \pi_i} = \frac{S_{1,i}}{\pi_i} - \nu \overset{\text{set}}{=} 0 \forall i, \sum_i \pi_i = 1 \quad \Rightarrow \quad \nu = 1, \pi_i = S_{1,i}.$$

Given the observation sequence $(Y_t)_t$ of an HMM, we can maximize the expected log likelihood under posterior probability measure using the Baum-Welch algorithm [Baum et al., 1970]:

$$\mathbb{E}_Q \overset{\text{def.}}{=} \mathbb{P}(\{S_t\}|\{Y_t\}) \, (\log \mathbb{P}(\{S_t, Y_t\}))$$

$$= \sum_i \mathbb{E}_Q (S_{1,i}) \cdot \log \pi_i + \sum_{t,i,j} \mathbb{E}_Q (S_{t,i} S_{t+1,j}) \cdot \log p_{i \to j} + \sum_{t,i} \mathbb{E}_Q (S_{t,i}) \log \mathbb{P}(Y_t|S_t = i),$$

where $\mathbb{E}_Q(S_{t,i})$ and $\mathbb{E}_Q(S_{t,i} S_{t+1,j})$ can be estimated through the forward-backward algorithm. By the same line of reasoning when the latent state sequence is known, the parameters that maximize the expected log likelihood are

$$p_{i \to j} = \sum_t \mathbb{E}_Q(S_{t,i} S_{t+1,j}) / \sum_t \mathbb{E}_Q(S_{t,i})$$

$$\pi_i = \mathbb{E}_Q(S_{1,i}).$$

Estimating the influence matrix and the inital marginal state distributionsfrom known state sequence $\left( S_t^{(c)} : c, t \right)$ follows the same procedure as in the Markov chain case. Recall that the log likelihood of a influence-model state sequence is

$$\log \mathbb{P} \left( \left\{ S_t^{(c)}, U_t^{(c)}, Y_t^{(c)} : c \in \{1, \cdots, C\}, t \in \{1, \cdots, T\} \right\} \right)$$

$$= \sum_{c=1}^{C} \sum_{i=1}^{m_c} S_{1,i}^{(c)} \log \pi_i^{(c)} + \sum_{t,c,c_1,i,j} U_t(c_1, c) S_{t,j}^{(c_1)} S_{t+1,i}^{(c)} \log H_{i,j}^{(c_1,c)} + \sum_{t,c,i} S_{t,i}^{(c)} \log \mathbb{P} \left( Y_t^{(c)} | S_t^{(c)} = i \right)$$

$$= \sum_{c,i=1}^{C,m_c} S_{1,i}^{(c)} \log \pi_i^{(c)} + \sum_{t,c,i=1}^{T,C,m_c} S_{t,i}^{(c)} \log \mathbb{P} \left( Y_t^{(c)} | S_t^{(c)} = i \right) +$$

$$+ \sum_{t,c,c_1,i,j=1}^{T-1,C,C,m_{c_1},m_c} U_t(c_1, c) S_{t,i}^{(c_1)} S_{t+1,j}^{(c)} \log A_{i,j}^{(c_1,c)} + \sum_{t,c,c_1=1}^{T-1,C,C} U_t(c_1, c) \log D^{(c_1,c)},$$

and the constraints are $\forall c, \sum_i \pi_i^{(c)} = 1$, $\forall c, \sum_{c_1} D^{(c_1,c)} = 1$, and $\forall c_1, c, i, \sum_j A_{i,j}^{(c_1,c)} = 1$. The Lagrangian is hence

$$\mathcal{L} = \log \mathbb{P} + \sum_{c_1,c,i} \mu_i^{(c_1,c)} (1 - \sum_j A_{i,j}^{(c_1,c)}) + \sum_c \nu^{(c)} (1 - \sum_i \pi_i^{(c)}) + \sum_c \omega^{(c)} (1 - \sum_{c_1} D^{(c_1,c)}).$$

We can get the maximum likelihood estimation of the the influence matrix and the marginal initial state distributions by setting the partial derivatives of the Lagrangian over the parameters to 0 and checking that the partial derivatives of the loglikelihood over the parameters are all negative.

$$\frac{\partial \mathcal{L}}{\partial A_{i,j}^{(c_1,c)}} = \sum_t U_t\left(c_1, c\right) S_{t,i}^{(c_1)} S_{t+1,j}^{(c)} / A_{i,j}^{(c_1,c)} - \mu_i^{(c_1,c)} \stackrel{\text{set}}{=} 0 \forall j, \sum_j A_{i,j}^{(c_1,c)} = 1$$

$$\Rightarrow \quad \mu_i^{(c_1,c)} = \sum_t U_t(c_1, c) S_{t,i}^{(c_1)}, A_{i,j}^{(c_1,c)} = \sum_t U_t\left(c_1, c\right) S_{t,i}^{(c_1)} S_{t+1,j}^{(c)} / \sum_t U_t\left(c_1, c\right) S_{t,i}^{(c_1)},$$

$$\frac{\partial \mathcal{L}}{\partial D^{(c_1,c)}} = \sum_{t=1}^{T} U_t\left(c_1, c\right) / D^{(c_1,c)} - \omega^{(c)} \stackrel{\text{set}}{=} 0 \forall c_1, \sum_{c_1} D^{(c_1,c)} = 1$$

$$\Rightarrow \quad \omega^{(c)} = T, D^{(c_1,c)} = \sum_t U_t\left(c_1, c\right) / T$$

$$\frac{\partial \mathcal{L}}{\partial \pi_i^{(c)}} = S_{1,i}^{(c)} / \pi_i^{(c)} - \nu \stackrel{\text{set}}{=} 0 \forall i, \sum_i \pi_i^{(c)} = 1$$

$$\Rightarrow \quad \nu = 1, \pi_i^{(c)} = S_{1,i}^{(c)},$$

Estimation the influence matrix and the initial marginal state distributions of a latent state influence model from an observation sequence $\left\{ Y_t^{(c)} : c, t \right\}$ follows the same procedure as in the HMM case. When the latent state sequence $\left\{ U_t^{(c)}, S_t^{(c)} : c, t \right\}$ is unknown, we can maximize the expected log likelihood under posterior probability measure

$$\mathbb{E}_{\mathbb{Q}} \stackrel{\text{def}}{=} \mathbb{P}\left(\left\{S_t^{(c)}, U_t^{(c)} : c, t\right\} | \left\{Y_t^{(c)} : c, t\right\}\right) \left(\log \mathbb{P}\left(\left\{S_t^{(c)}, U_t^{(c)}, Y_t^{(c)} : c, t\right\}\right)\right)$$

$$= \sum_{c,i=1}^{C,m_c} \mathbb{E}_{\mathbb{Q}}\left(S_{1,i}^{(c)}\right) \cdot \log \pi_i^{(c)} + \sum_{t,c,i=1}^{T,C,m_c} \mathbb{E}_{\mathbb{Q}}\left(S_{t,i}^{(c)}\right) \cdot \log \mathbb{P}\left(Y_t^{(c)} | S_t^{(c)} = i\right) +$$

$$+ \sum_{t,c,c_1,i,j=1}^{T-1,C,C,m_{c_1},m_c} \mathbb{E}_{\mathbb{Q}}\left(U_t(c_1, c)S_{t,i}^{(c_1)} S_{t+1,j}^{(c)}\right) \cdot \log A_{i,j}^{(c_1,c)} + \sum_{t,c,c_1=1}^{T-1,C,C} \mathbb{E}_{\mathbb{Q}}\left(U_t(c_1, c)\right) \cdot \log D^{(c_1,c)},$$

where $\mathbb{E}_{\mathbb{Q}}(U_t(c_1, c)S_{t,i}^{(c)})$ and $\mathbb{E}_{\mathbb{Q}}(U_t(c_1, c) \cdot S_{t,i}^{(c_1)} S_{t+1,j}^{(c)})$ can be estimated. By the same line of reasoning when the latent state sequence is known, the parameters that maximize the expected log likelihood are

$$A_{i,j}^{(c_1,c)} = \sum_t \mathbb{E}_{\mathbb{Q}}\left(U_t\left(c_1, c\right) S_{t,i}^{(c_1)} S_{t+1,j}^{(c)}\right) \bigg/ \sum_t \mathbb{E}_{\mathbb{Q}}\left(U_t\left(c_1, c\right) S_{t,i}^{(c_1)}\right),$$

$$D^{(c_1,c)} = \sum_t \mathbb{E}_{\mathbb{Q}}\left(U_t\left(c_1, c\right)\right) / T$$

$$\pi_i^{(c)} = \mathbb{E}_{\mathbb{Q}}\left(S_{1,i}^{(c)}\right).$$

Sometimes we want to compute the influence matrix from either the influence process $\left\{ S_t^{(c)} : c, t \right\}$ or the posterior probability distribution of this process given observations, and we do not want to

make any assumptions about the unknown "influencer" sequence $\left\{ U_t^{(c)} : c, t \right\}$ such as an assumption about the probability distribution. When we know the latetent state sequence $\left\{ S_t^{(c)} : c, t \right\}$, the computation of the influence matrix doesn't require enumerating the exploding number of latent state combinations of the interacting processes $\left\{ S_t^{(c)} : c \right\}$ and is hence tractable. In contrast, when we only know some statistics about the latent state sequence, the computation normally involves an exploding number of terms that express the different influences corresponding to different latent state combinations $\left\{ S_t^{(c)} : c \right\}$, and I only know some approximation algorithms that combine the exploding number of terms into a computationally manipulatable number of terms.

Computing the influence matrix that maximizes the likelihood of a known latent state sequence $\left\{ S_t^{(c)} : c, t \right\}$ corresponds to the following optimization problem:

$$
\underset{a,d}{\operatorname{argmax}} \underbrace{\sum_{c=1}^{C} \pi_{S_1^{(c)}}^{(c)} + \sum_t \sum_c \log \sum_{c'} a_{S_t^{(c')} S_{t+1}^{(c)}}^{(c',c)} d^{(c',c)}}_{\log \mathbb{P}\left(\left\{ S_t^{(c)} : c, t \right\}\right)}
$$

$$
\text{satisfying } \forall c, \sum_{c'=1}^{C} d^{(c',c)} = 1; \forall c', c, i, \sum_j a_{i,j}^{(c',c)} = 1; \forall c, \sum_{i=1}^{m_c} \pi_i^{(c)} = 1.
$$

This maxmization problem is benign because the log-likelihood to be maximized is an increasing function of the variables $a$ and $d$, the log-likelihood takes as its domain the convex hull of $a$ and $d$, and when the latent influence process is aperiodic and non-reducible the log-likelihood function takes as its domain the interior of convex hull. The partial derivatives of the log-likelihood over $a$ and $d$ are respectively

$$
\frac{\partial \log \mathbb{P}}{\partial d^{(c_1,c)}} = \sum_t \left( a_{S_t^{(c_1)} S_{t+1}^{(c)}}^{(c_1,c)} \Big/ \sum_{c'} a_{S_t^{(c')} S_{t+1}^{(c)}}^{(c',c)} d^{(c',c)} \right)
$$

$$
\frac{\partial \log \mathbb{P}}{\partial a_{i,j}^{(c_1,c)}} = \sum_{\left\{ t : S_t^{(c_1)} = i, S_{t+1}^{(c)} = j \right\}} \left( a_{i,j}^{(c_1,c)} \Big/ \sum_{c'} a_{S_t^{(c')} S_{t+1}^{(c)}}^{(c',c)} d^{(c',c)} \right).
$$

## 2.5 Example: Interacting Processes with Noisy Observations

Let us suppose that we have six stochastic processes, and we sample these six processes with six sensors. Each process can be either signaled (one) or non-signaled (zero) at any time, and the corresponding sensor has approximately 10% of its samples flipped. The interaction of the six stochastic processes behind the scene looks like this: processes one through three tend to have the same states; processes four through six tend to have the same states; the processes are more likely to be non-signaled than to be signaled; and the processes tend to stick to their states for a stretch of time. The parameters of the model are given as the following and are going to be estimated:
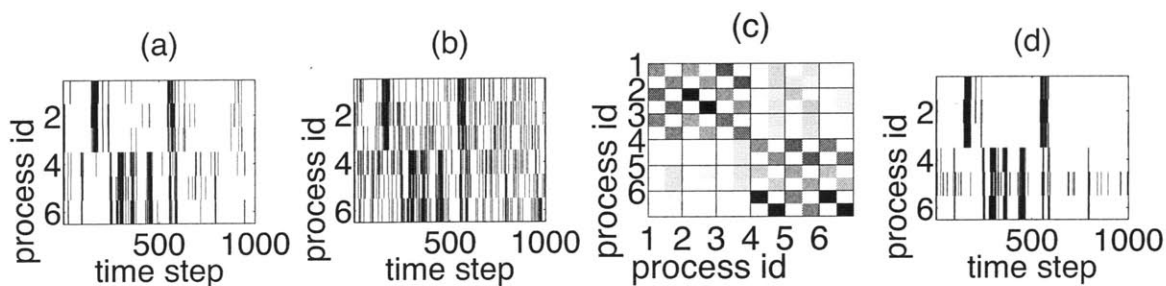
Figure 2.3: Inference from observations of interacting dynamic processes.

$$A_{ij} = \begin{pmatrix} .99 & .01 \\ .08 & .92 \end{pmatrix}, 1 \leq i, j \leq 6, \; B_i = \begin{pmatrix} .9 & .1 \\ .1 & .9 \end{pmatrix}, 1 \leq i \leq 6, \; d_{ij} = .33, 1 \leq i, j \leq 3, \text{ and}$$
$$d_{ij} = .33, 4 \leq i, j \leq 6.$$

In Figure 2.3, (a) shows the sampled latent state sequences, (b) shows the corresponding observation sequences, (c) shows the influence matrix reconstructed from sampled observation sequences, and (d) shows the reconstructed latent state sequences after 300 observations. The $(i, j)^{th}$ entry of the $(c_1, c_2)^{th}$ sub-matrix of an influence matrix determines how likely that process $c_1$ is in state $i$ at time $t$ and process $c_2$ is in state $j$ at time $t + 1$. It can be seen from Figure 2.3 (c) that the influence model computation recovers the structure of the interaction.

The influence model can normally attain around $95\%$ accuracy in predicting the latent states for each process. The reconstructed influence matrix has only $9\%$ relative differences with the original one. Using only observations of other chains we can predict a missing chain's state with $87\%$ accuracy.

We then constructed a more complex experimental setting to compare the performances of different types of hidden Markov models. In this setting, we have a Markov process with $2^C$, where $C = 10$, number of states and a randomly generated state transition matrix. Each system state $\vec{s}_t$ is encoded into a binary $s_t^{(1)} \cdots s_t^{(C)}$. Each of the $m_c = 2$ evaluations of "bit" $s_t^{(c)}$ corresponds a different 1-d Gaussian observation $o_t^{(c)}$: Digit $s_t^{(c)} = 1$ corresponds to $o_t^{(c)} \sim \mathcal{N}[\mu_1 = 0, \sigma_1^2 = 1]$ ; Digit $s_t^{(c)} = 2$ corresponds to $o_t^{(c)} \sim \mathcal{N}[\mu_2 = 1, \sigma_2^2 = 1]$ . Figure 2.4 compares the performances of several dynamic latent structure models applicable to multi-sensor systems. Of the 1000 samples $(\vec{o}_t)_{1 \leq t \leq 100}$ , we use the first 250 for training and all 1000 for validation.

There are two interesting points. First, the logarithmically scaled number of parameters of the influence model allows us to attain high accuracy based on a relatively small number of observations. This is because the eigenvectors of the master Markov model we want to approximate are either mapped to the eigenvectors of the corresponding influence model, or mapped to the null space of the corresponding event matrix thus is not observable from the influence model, and that in addition the eigenvector with the largest eigenvalue (i.e., 1) is mapped to the eigenvector with the largest eigenvalue of the influence matrix [Asavathiratham, 1996]. Secondly, both the influence model and the hidden Markov model applied to individual processes are relatively immune to over-fitting, at the cost of low convergence rates. This situation is intuitively the same as the numerical analysis wisdom that a faster algorithm is more likely to converge to a local extremum or to diverge.
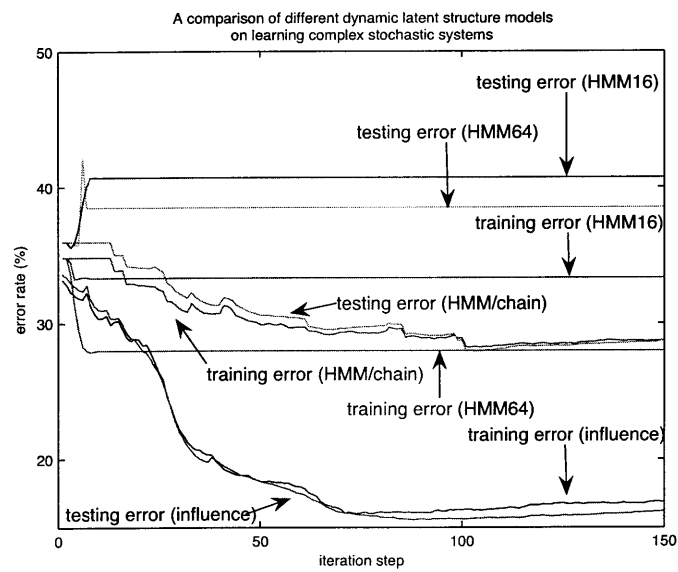
Figure 2.4: Comparison of dynamic models.

# Chapter 3

# Modeling and Optimizing Group Collaboration Dynamics

An understanding of face-to-face communications within a group discussion can provide new clues about how humans collaborate to accomplish complex tasks and how the collaboration protocols can be learned. It can also help evaluate and facilitate brainstorming sessions. I will discuss the following three findings about the dynamics: (a) meetings in different languages and on different topics could follow the same form of dynamics; (b) the functional roles of the meeting participants could be better understood by examining not only their individual speaking and activity features but also their interactions with each other; (c) the outcome of a meeting could be predicted by examining how its participants interact.

The *form* of the group discussion process (in which participants present their opinions, argue about an issue, and try to reach a consensus), is related to the participants' tendencies to maximize the outcome of the discussion. People "instinctively" know how to cope with each other in many different situations to make an effective discussion. As a result, we could expect some invariant structures from one discussion to another, and could, just by watching the discussion dynamics, answer the following four questions: (1) how the opinions of the meeting participants differ from each other, (2) what are the psychological profiles of the participants, (3) how the discussion progress, and (4) whether the discussion is effective. We do not concern ourselves with the content of the discussion, and thus we cannot answer questions concerning the content., Even so, we can nevertheless learn a great deal simply by observing the form (i.e., the container).

I will model the group discussion dynamics as interacting stochastic processes, with each process representing a participant. I will identify the different functional roles the participants take at each point in a group discussion, and evaluate the discussion efficiency within the framework of the stochastic process. I will first briefly discuss the intuition behind the influence modeling of the group interaction process, the mission survival data set, and the previous work related to understanding group dynamics, influences, meeting progression, and the effectiveness of a meeting. The study of non-verbal aspects of a face-to-face group discussion is not new, yet the approach used here yields better accuracy when estimating the participants' functional roles and the discussion outcome than obtained in previous studies. This is because this new formulation takes into account the interaction features. I will then describe several data sets and give the new results on their inter-

action statistics. The statistics both motivate our new formulation of *influence* and provide insights about why the new formulation could give better estimation results. I will proceed to compare the old and new formulations on *influence* and give the performance of estimating the functional roles with influence modeling and other methods. I will conclude by briefly describing the experiences and lessons I have learned in my efforts to understand the non-verbal aspects of group discussion, as well as offer some suggestions for future research .

## 3.1   Influence Modeling of Interactive Process Dynamics

Let us employ a thought experiment to examine an imaginary group discussion which involves the participants expressing their opinions about an open problem in order to reach a consensus. The quantitative analysis corresponding to the following description will be given in the following sections.

A discussion is normally driven by one person at a time, and it can be driven by different persons at different times. The person who drives the discussion normally has longer uninterrupted speaking time, steadier intonation and speaking speed. They also usually receive more attention from the other persons, who usually display their attention by orienting their bodies toward and looking at the speaker.

The persons who do not drive the discussion (listeners) can, from time to time and individually, support the person who drives the discussion (turn-taker) by briefly showing their agreement or by briefly adding supporting material to the turn-taker's argument. The listeners can sometimes request clarifications on the turn-taker's argument, and the requested clarification can subsequently be provided by either the turn-taker or the other listeners.

One or more listeners may infrequently show their disagreement with the turn-taker's opinion/argument and initiate an "attack," which may consequently pull more listeners into the "battle." The intensity of the battle is indicated by significantly less body/hand movement by the person who initiates the attack, significantly more body/hand movement of the others in response to the attack-initiator (who speak and turn to each other), and the large number of simultaneous speakers.

The turn-taker and a listener may, from time to time, engage in a series of back-and-forth negotiations to fill the gap in their understandings or opinions. If the negotiation takes too long, the other participants may jump in and terminate the negotiation. When the turn-taker finishes his turn, he may either simply stop speaking or explicitly hand over his turn to a listener.The next turn-taker will continue to drive the discussion appropriately.

In many of the discussions, there is a distinctive "orienteer," who has the "charisma" to drive the discussion forward when it comes into a halt or degenerates into chaos. This charisma is reflected by the capability of the orienteer to quickly seize the attention of the others. When the orienteer takes on the orientation role, all other speakers quickly turn their body towards the orienteer, and the other current (normally multiple) speakers quickly stop speaking.

What has just been described can be expressed as an "influence model," in which each participant randomly chooses to maintain his role (e.g., turn-taker, supporter, attacker, orienteer) for some duration or chooses to make a transition to another role. The duration for which a turn-taker drives

the discussion depends among many factors, including the option and arguments of the turn-taker, his own style, and the responses of the other participants. When one person expresses his opinion and arguments, the other people normally listen to his statement attentively and patiently, and express their agreements and doubts unobtrusively. The transition from one turn-taker to another depends on how the latter's opinion is related to the former's, and how the latter wants to drive the discussion. The manner and the likelihood for a participant to express his support, doubt or disagreement depend both upon his judgment about the importance of making an utterance and upon his personal style.

An observer who watches the group discussion dynamics — the turn-takers at each point, the transitions of the turns, the responses of the listeners and the dyadic back-and-forth negotiations — can often obtain a precise understanding of them (the testing data set) by pattern-matching them with the past group discussion dynamics (the training data set) stored in his memory. The reason for is because the multi-person face-to-face interactions normally take on a small number of regular patterns out of a huge number of possibilities. For example, if each of the four participants in a discussion can take one of the four roles — protagonist, attacker, supporter or neutral, there will be $4^4 = 256$ possible role combinations. However, in an efficient group discussion, only a few combinations exist most of the time. A corollary of the regularity of face-to-face interactions is that we can evaluate the effectiveness of a group discussion by examining how frequently its interaction pattern is an efficient one.

Since the dynamics of a discussion is dependent upon the purpose of the discussion, we can either imagine the characteristics of an efficient discussion for a certain purpose, or compute the characteristics based on simplified mathematical models representative of the discussion purpose. We can also use our intuition either to guide our experimental designs or to help interpreting the experimental results. Since we normally obtain a more comprehensive range of perspectives by listening to more people during the discussion of an open problem, and since we can normally only pay attention to a single person at a time, we could imagine that, for the most part, an effective discussion is driven by a single person. Since discussing a topic normally requires a considerable amount of set-up time and summarizing time, we should not see frequent transitions among topics. The topics can be separated from each other by different amounts of participation by the individuals and different interaction dynamics. Since a back-and-forth dyadic negotiation normally involves the interests and attention of only two individuals, it should generally not last long in an effective group discussion. Thus, the effectiveness of a group discussion could be studied using stochastic process models and statistical learning methods.

Different group-discussion purposes require different types of dynamics. Even so, there are invariants in interpersonal communications. The cognitive loads of individuals has a statistical distribution. Different types of turn-taking dynamics statistically result in different performances, conditioned by the meeting purposes and the individual parameters. Similar kinds of group-discussion issues, such as blocking and social loafing [Diehl and Stroebe, 1991, Jackson and Harkins, 1985] may exist in different types of discussions. As a result, while the same stochastic model may be used to fit all group-discussions, different purposes may require different parameters. We should take special care of the compatibility of two group discussions when we fit a dynamic model to the former with appropriate parameters and apply the fitted model to the latter.

## 3.2   Mission Survival Corpus I

The Mission Survival Corpus [Pianesi et al., 2008b] is a multimodal annotated corpus based on the audio and video recordings of eight meetings that took place in a laboratory setting equipped with cameras and microphones. Each meeting consisted of four people attempt to solve the "mission survival task." This task is frequently used in experimental and social psychology studies to elicit decision-making processes in small groups. Originally designed by National Aeronautics and Space Administration (NASA) to train astronauts, the Survival Task proved to be a good indicator of group decision making processes [Hall and Watson, 1970]. The exercise promotes group discussion by asking participants to reach a consensus about how to survive in a disaster scenario, such as a moon landing or a plane crash in Canada. The group is required to rank a number of items (usually 15) according to their importance for the survival of crew members.- For this study setting, the plane crash version was used. This consensus decision-making scenario was chosen for the purpose of meeting dynamics analysis, primarily because of the intensive engagement required of groups in order to reach a mutual agreement. This offered the possibility to observe a large set of social dynamics and attitudes. In this setting, the basic structure of the Survival Task was retained with minor adjustments: (a) the task was competitive across groups/team, with a prize being awarded to the group providing the best survival kit; (b) the task was collaborative and based upon consensus within the group, meaning that a participant's proposal became part of the common sorted list only if he/she managed to convince the other of the validity of his/her proposal.

The recording equipment consisted of five FireWire cameras.Four were placed at the four corners of the room, and one was placed directly above the table. Four web cameras were also installed on the walls surrounding the table. Speech activity was recorded using four close-talk microphones, six tabletop microphones and seven T-shaped microphone arrays, each consisting of four omni-directional microphones installed on the four walls. This configuration was used in order to obtain an optimal coverage of the environment for speaker localization and tracking. Each session was automatically segmented, labeling the speech activity recorded by the close-talk microphones every 330ms [Carli and Gretter, 1997]. The fidgeting—the amount of energy in a person's body and hands—was automatically tracked by using skin region features and temporal motion [Chippendale, 2006]. The values of fidgeting for hands and body were collected for each participant and normalized on the fidgeting activity of the person during the entire meeting.

The Functional Role Coding Scheme (FRCS) was partially inspired by Bales' Interaction Process Analysis [Bales, 1969]. This consists of ten labels that identify the behavior of each participant in two complementary areas. The first was referred to as the Task Area, which includes functional roles related to facilitation and coordination tasks as well as to technical experience of members. The second was the Socio-Emotional Area, which was concerned with the relationships between group members and the functional roles oriented toward the functioning of the group as a group. A synthetic description of the FRCS is given below (for more information, see [Pianesi et al., 2008b]. The Task Area functional roles consisted of the following: the orienteer (o), who oriented the group by introducing the agenda, defined goals and procedures, kept the group focused and summarized the most important arguments and the group decisions; the Giver (g), who provided factual information and answers to questions, stated her beliefs and attitudes about an idea, and expressed personal values and factual information; the Seeker (s), who requested information, as well as clarifications, to promote effective group decisions; the Procedural Technician (pt), who

used the resources available to the group, managing them for the sake of the group; the follower (f), who merely listened, without actively participating in the interaction. The Socio-Emotional functional roles consisted of: the Attacker (a), who deflated the status of others, expressed disapproval, and attacked the group or the problem; the Gate-keeper (gk), who was the group moderator, mediated the communicative relations, encouraged and facilitated the participation and regulated the flow of communication; the Protagonist (p), who took the floor, drove the conversation, assumed a personal perspective and asserted her authority; the Supporter (su), who displayed a cooperative attitude demonstrating understanding, attention and acceptance as well as providing technical and relational support; the Neutral Role (n), played by those who passively accept the ideas of the others, serving as an audience in group discussion. Of course, participants may—and often do—play different roles during the meeting. However at any given time, each of them plays exactly one role in the Task Area and one role in the Socio-Emotional Area. The FCRS was shown to have a high inter-rater reliability (Cohen's statistics $\kappa = 0.70$ for the Task Area; $\kappa = 0.60$ for the Socio-Emotional Area).

## 3.3 A Literature Review on the Form and Content of Interactive Group Problem Solution

Various approaches have been applied to characterize the roles of speakers in news bulletins, based on the distinctive characteristics of those roles. Vinciarelli [Vinciarelli, 2007b,a] used Bayesian methods to identify the roles of the anchorman, the second anchorman, the guest as well as several other roles. These roles were assigned based upon how much each person speaks, when they speak (beginning, middle, or end of the bulletin), and the sequence in which they speak. Using material from three movies, based on the co-occurrences of roles in different scenes, the same social network analysis (SNA) concept was adopted by Weng et al. [Weng et al., 2007] to identify the hero, the heroine, and their respective friends. When analyzing a radio program, Barzilay et al. [Barzilay et al., 2000] made use of the keywords employed by individuals , the durations of their speaking turns, and the explicit speaker introduction segments to identify the roles of the anchor, the journalists and the guest speakers.

Different meeting states and roles have been defined, and their characteristics and estimation algorithms have been studied. Banerjee and Rudnicky [Banerjee and Rudnicky, 2004] defined three meeting states (discussion, presentation and briefing) and four corresponding roles (discussion participators, presenter, information provider, and information giver). They subsequently used the C4.5 algorithm to describe the meeting states and the roles, based on four features (number of speaker turns, number of participants who spoke, number of overlaps, and average length of overlaps). McCowan et al. [McCowan et al., 2005] developed a statistical framework based on different Hidden Markov Models. This enabled them to recognize the sequences of group actions starting from audio-visual features concerning individuals' activities—e.g., "discussion," as a group action recognizable from the verbal activity of individuals. In a simulated discussion on the development of a new remote control, Garg et al. [Garg et al., 2008] discussed the recognition of the project manager, the marketing expert, the user interface expert and the industrial designer. . The role of recognizer is based upon when the participants speak and what keywords the participants use.

The subject of "Dominance detection" has aroused a great deal of interest. This is probably because the dominant person is believed to have large influence on the outcome of a meeting. Rienks et al. [Rienks and Heylen, 2005, Rienks et al., 2006] used various static and temporal models to estimate the dominance of the participants in a meeting, and concluded that the automated estimation is compatible with the human estimation. They used several nonverbal features, including speaker turns, floor grabs, speaking length. They also examinedverbal features, such the, number of spoken words used by the next speaker, as well as audio features retrieved from the transcript of the discussion. Jayagopi et al. [Jayagopi et al., 2008, Hung et al., 2008b,a] extended the work of Rienks et al., and estimated dominance using features computed from the audio and video recordings — e.g., total speaking "energy," total number of times being unsuccessfully interrupted.

The social psychology literature, especially that related to the structures and the performances of small group discussions, provides useful observations, insights, and challenges to investigate using automated computer algorithms. Conversation and discourse analysis provides useful observations and examples [Sacks et al., 1974, Atkinson, 1985, Frey et al., 1999, Schegloff, 2007]. The features and structures of conversational group processes may be analyzed through experiments and simulations. Bales investigated the phases (e.g., giving opinion, showing disagreements, asking for suggestion) and the performances of group discussions, as well as the different roles that the discussion participants play [Bales, 1950, 1969, Bales and Cohen, 1979]. In contrast, McGrath examined meetings based on their different tasks [McGrath and Kravitz, 1982, McGrath, 1984]. The usefulness of group brainstorming has been widely debated [Osborn, 1963, Frey et al., 1999] Production blocking and social loafing have been identified as two drawbacks of group brainstorming [Diehl and Stroebe, 1991, Karau and Williams, 1993, Nijstad et al., 2003]. Hall [Hall and Watson, 1970] and Wilson [Wilson et al., 2004] systematically analyzed their respective group brainstorming experiments, and have each offered explanation for why a group can collectively outperform the individuals within it.

The work related to the Mission Survival corpus includes several distinct components. First of all, it requires the identification of functional relational roles (social and task roles). This issue was addressed by Zancanaro et al. [Zancanaro et al., 2006, Pianesi et al., 2008b] through an SVM that exploited speech activity (whether a participant was speaking at a given time) and the fidgeting of each participant in a time window. Dong [Dong et al., 2007] extended this work by comparing the SVM-based approach to HMM-based and IM-based approaches. Pianesi et al. [Pianesi et al., 2008a] have used social behavior as a means to identify individual characteristics, such as personality traits. The task consisted of a three-way classification (low, medium, high) of the participants' levels of extroversion and locus of control. The study also used speech features that were demonstrated to be "honest signals" for social behavior, as well as visual fidgeting features.

As far as can be determined, this present study is the first to discuss the features and modeling issues of the turn-taking behavior and the personal styles in an unconstrained group discussion that can be extracted with computer algorithms from the audio and video recordings. This study also presents the initial findings on the correlation between discussion turn-taking behavior and discussion performance. The discussion is based on *Mission Survival Corpus I*. The primary difficulty in the current work is that was studying an unconstrained group discussion. Thus, in contrast to communications such as news bulletins, there was no any pre-defined agenda or keywords would could be used, , nor was there any visual cues, such as a whiteboard or a projector screen. A person who was dominant in one part of a discussion may be non-dominant in another part. Nevertheless,

the study revealsthat, although the predefined macro-structure does not exist in an unconstrained discussion, the micro-structures at different parts of the discussion are based on the instantaneous roles of the participants, and the statistics associated with the micro-structures are related to the discussion performance.

The influence modeling used in this work has a long history of development, and simultaneously captures interactions and temporal coherence The coupled hidden Markov models were first the development to capture the interactions and temporal coherence of two parts based on audio and visual features [Brand et al., 1997, Brand, 1996, Oliver et al., 2000]. Asavathiratham introduced the influence model to study the asymptotic behavior of a large number of individual power plants in a network [Asavathiratham, 1996, Asavathiratham et al., 2001]. The approximation used by Asavathiratham is that the probability measure of state of a power plant is a linear functional of the probability measures of all power plants' states in the network. A similar concept was employed by Saul and Boyen [Saul and Jordan, 1999, Boyen and Koller, 1999]. Choudhurry noted that individuals have their characteristic styles in two-person face-to-face conversations, and the overall style of a two-person face-to-face conversation more closely ressembles the style of the more influential person. Choudhurry et al. subsequently used influence modeling to study the structures of discussions and organizations [Basu et al., 2001, Choudhury and Pentland, 2002, Choudhury, 2003]. Dong developed several versions of multi-agent dynamic Bayesian networks using the same name, which are better fitted with the probability measures of group processes [Dong, 2006, Dong and Pentland, 2007].

## 3.4 Group Process Statistics in *Mission Survival Corpus I*

Each individual within a group discussion has their own characteristic style with regards to the frequency, the duration and the functional (i.e., task and socio-emotional) roles of their speaking turns. In *Mission Survival Corpus I*, some individuals consistently take certain functional roles, , while other individuals display little or no consistency in the roles they take. . The functional roles each have their respective characteristics, durations , and interactions with other functional roles, independently of who assumes them. As a result, the functional roles of a speaker turn can be inferred from the characteristics of the turn and the characteristics of the turn-taker.

Figure 3.1 displays the decision trees that reveal a meeting participant's functional roles at a specified moment, as a function of the amounts of time he speaks in the time windows of different sizes around the moment. The C4.5 algorithm is used to generate the decision trees from four discussions of *Mission Survival Corpus I* as training data.It correctly captures the characteristics of the functional roles. An information giver speaks more than an information seeker in a short time window. A protagonist speaks more than a supporter in a long time window. A neutral role (i.e., a listener or a follower) speaks much less than the other roles in time windows of up to several minutes. The C4.5 algorithm, like many other modern statistical learning algorithms, is protected against overfitting by a mechanism. The trained decisions trees can attain an accuracy of approximately 55%. (As a comparison, the inter-rater reliability has Cohen's statistics $k = .70$ for the Task Area and $\kappa = 0.60$ for the Socio-Emotional area.) Further accuracy can be achieved by considering the speaker characteristics and more functional role characteristics. Since the participant who spends more time giving information often spends more time in seeking information ($R^2 = .27, F = 12.4$

(a) Task roles.                    (b) Socio-Emotional roles.
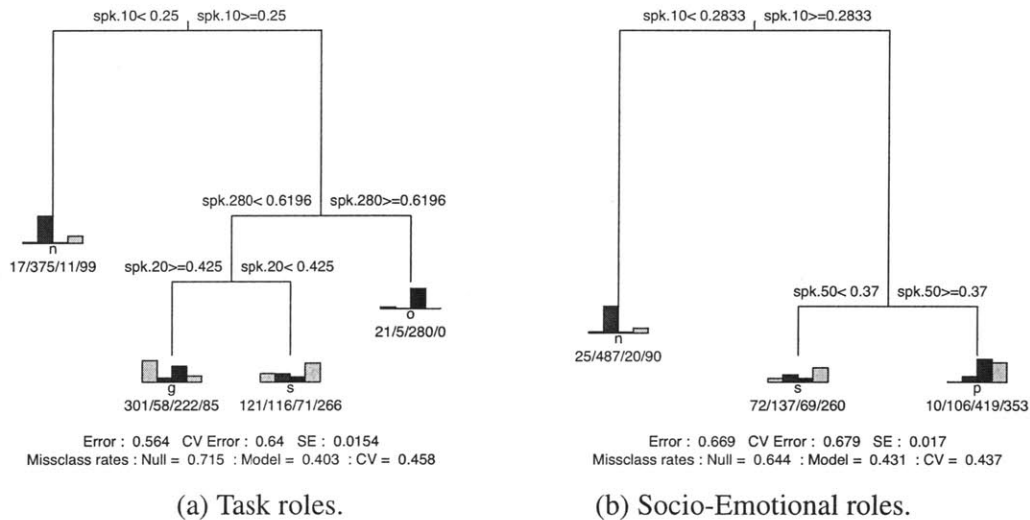
Figure 3.1: Decision trees trained with the C4.5 algorithm for functional role detection. If a person takes the Neutral/Follower role at a moment, he speaks noticeably less in the 10-second window around the moment. The Giver speaks more than the Seeker in a 20-second window. The Protagonist speaks more than the Supporter in a 50-second window. The Orienteer, on average, speaks 62% of time in a 280-second window. The speaking/non-speaking signal seems to be insufficient to detect an Attacker.

on 1 and 30 degrees of freedom, $p = .0014$), the total amount of time a participant has spent in giving information can be used to determine whether a short speaking turn corresponds to a seeker role or a neutral role. Due to the way an auxiliary role, such as seeker/supporter, and a major role, such as giver/protagonist co-occur, the amounts of speaking time of a participant in time windows of different sizes can be contrasted with those of the other participants to disambiguate the roles of the participant. Since an attacker is relatively quiet and arouses significant agitation from the others, and since a person in a neutral role is often less paid attention, the intensities of hand/body movements can be taken as the characteristics of those roles.

This section is organized into two subsections. In Subsection 3.4.1, the durations of each functional role and the likelihood that different functional roles co-occur is analyzed. In Subsection 3.4.2, the issue of who is more likely to take which roles, based ontheir individual honest signals, is analyzed.

## 3.4.1   Turn-taking behavior

The patterns in the functional roles, social signaling, turn-taking behavior, and their relations are given as follows. Any effective heuristics and statistical learning methods that model the group discussion behavior should take advantage of these patterns.

We will first examine the (Task Area and Socio-Emotional Area) role assignments of the subjects in *Mission Survival Corpus I*. The role assignments reflect how the observers understand the group

processes.

Table 3.1 (left) gives the durations in seconds of social roles, task roles and their combinations. In this table, an instance of a supporter role has a significantly shorter average duration than that of a protagonist role (15 vs. 26). This coincides with the fact that a protagonist is the main role which drives the conversation while a supporter takes a secondary importance. An attacker role takes an average duration of 9 seconds, which is equivalent to 10~20 words and approximately one sentence, assuming conversations occur at a rate of approximately 100 - 200 words per minute. This reflects a person's strategy for expressing his contrasting ideas concisely, so that he can make constructive utterances and simultaneously avoid conflicts. A person asks questions (when he takes an information-seeker's role) more briefly than when he provides information (when he takes an information-giver's role). This reflects the natural tendency to make task-oriented discussions more information-rich and productive. A protagonist role is, on average, 37% longer in duration. This indicates that the social roles happen on a different time scale compared with the task roles. The two types of roles are not perfectly correlated with each other. A discussion is normally driven by one person and thus generally has only a single protagonist at any given point in time. The protagonist can ask another person questions, and the latter generally gives the requested information briefly, avoiding assuming the speaker's role for too long. The protagonist can seldom be interrupted by questions, and the questioner generally seeks additional information in a brief and collaborative way. The durations of the neutral roles in the "task role dimension" and the "social role dimension" are less than twice the durations of the giver's role and the protagonist's role, respectively. This indicates the participants do not passively listen when they take listeners' roles.

Table 3.1 (right) shows the number of speaking-turns and the amount of time that the meeting participants take different task roles, social roles and combinations of task and social roles. This table complements Table 3.1 (left) regarding how individual participants take different functional roles in a discussion. The group process can also be viewed as a Markov process with different distributions of functional roles at different time. In *Mission Survival Corpus I*, the configuration *1g3n0o0s*, which denotes the configuration of the discussion with 1 Giver, 3 Neutrals, 0 Orienteer, and 0 Seekers, takes 36% of discussion time.The configurations *2g2n0o0s*, *0g3n1o0s*, *0g4n0o0s*, *1g2n0o1s*, *3g1n0o0s*, *1g2n1o0s*, *0g3n0o1s* and *2g1n0o1s* take 20%, 13%, 11%, 5%, 5%, 4%, 2% and 1% discussion time respectively. In the same data set, the different socio-emotional role distributions *0a3n1p0s*, *0a4n0p0s*, *0a3n0p1s*, *0a2n2p0s*, *0a2n1p1s*, *0a2n0p2s*, *0a1n2p1s*, *0a1n3p0s* and *0a1n1p2s* take 36%, 21%, 18%, 11%, 7%, 3%, 1%, 1% and 1% of discussion time, respectively. Each different role distribution tends to last for a specific duration. Considering a group process in terms of role distribution makes the group process model speaker-independent, and thus effectively compresses the number of states of the group process. Since it is possible to model a group process more effectively in terms of "influence," we will not discuss the distribution of roles further.

We will proceed to analyze the turn-taking behavior, the body movements, and the hand movements corresponding to different roles. The analysis will show that, rather than being artificially imposed to the group processes, the roles reflect a set of essential features of the group processes.

In a discussion involving multiple persons, the individuals normally orient their bodies towards the locus of the discussion, which can be either the protagonist or the information giver. The

Table 3.1: (Left) Durations in seconds of social roles , task roles and their combinations. (Right) Number of instances and amount of time in seconds that a person takes a task role, a social-role and a task/social role combo. Each table entry $\mu(\sigma)$ gives the mean and standard deviation for a specific case.

| $\mu(\sigma)$ | a | n | p | s | marginal |
|---|---|---|---|---|---|
| g | 8(6) | 10(16) | 23(24) | 11(7) | **19(20)** |
| n | 2(2) | 52(79) | 4(6) | 5(5) | 34(45) |
| o | N/A | 4(8) | 10(9) | 18(16) | 17(14) |
| s | 7(4) | 6(4) | 9(7) | 10(5) | **9(5)** |
| marginal | 9(4) | 56(85) | 26(27) | 15(14) | 7(50) |

| | a | n | p | s | total |
|---|---|---|---|---|---|
| g | 5(39) | 316(3k) | 233(5k) | 112(1k) | 666(9k) |
| n | 9(22) | 426(22k) | 185(747) | 147(718) | 767(24k) |
| o | 0(0) | 67(323) | 21(432) | 53(1k) | 141(2k) |
| s | 5(36) | 74(471) | 27(253) | 17(170) | 123(930) |
| total | 19(97) | 883(26k) | 466(6k) | 329(3k) | 1697(36k) |

Table 3.2: Average percentage of speaking time in 10-second windows (spk) around different social roles and task roles, average body movement (bdy) and hand movement (hnd) of the self (self) and the others (othr) in the 10-second windows around the shifts into different social roles. Each table entry $\mu(\sigma)$ gives the mean and standard deviation for a feature-role combination.

| $\mu(\sigma)$ | | Social-Emotional Area Roles | | | | Task Area roles | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Attacker | Neutral | Protagonist | Supporter | Giver | Follower | Orienteer | Seeker |
| features | self spk | .47(.16) | .20(.21) | .62(.21) | .54(.21) | .58(.21) | .16(.18) | **.62(.21)** | .41(.19) |
| | othr spk | .30(.09) | .32(.12) | .25(.14) | .25(.15) | .25(.14) | .33(.12) | **.23(.14)** | .30(.12) |
| | self hnd | 11(14) | 18(21) | 18(21) | 16(19) | 17(21) | 18(20) | 17(20) | 15(17) |
| | othr hnd | 20(14) | 16(13) | 19(14) | 17(12) | 19(14) | 16(13) | 18(12) | 18(14) |
| | self bdy | **11(19)** | 20(22) | 21(22) | 18(20) | 19(22) | 20(22) | 20(20) | 18(21) |
| | othr bdy | **23(14)** | 19(14) | 22(14) | 19(14) | 21(14) | 18(13) | 21(14) | 19(14) |

protagonist or the information giver normally transmits non-verbal communications to the listeners by turning his body toward them. The shifts of attention consist of a significant fraction of hand and body movementa. In the mission survival data set, the correlation between the change-of-speaker and the body/hand movement intensity is greater than 0.50.

Table 3.2 shows how the meeting participants execute their Task Area Roles and Socio-Emotional Area Roles in terms of \the amount they speak and to whom attentions is given. While the patterns are weak and might not be sufficient for constructing good role classifiers, they nevertheless exist and match our intuition: An attacker provokes a significant amount of attention, hand movements and body movements from the others, while he himself shows significant less hand and body movements. The neutral roles, on average, the supporter role and the seeker role attract less attention from the others compared with the giver role and the protagonist role. In the 10-second window when a person takes either a supporter role or a seeker role, he displays less hand and body movements. This may be due to the fact he has already paid attention to the locus of the discussion when he takes one of those roles. When a person takes an Orienteer role, on average, only 23% of the time in the 10-second window do the other three participants speak. The Orienteer speaks 62% of the time in this window. This indicates that the one task of an Orienteer is to keep the brainstorming on track.

Table 3.3 shows how the meeting participants shift their roles as a function of speaker overlap. We intuitively view speaker overlap as an indicator of the intensivity of a discussion. The tables

Table 3.3: Distribution of social roles and task roles conditioned on the amount of speaker-overlap.

| | | Socio-Emotional Area Roles | | | | $\Sigma$ | Task Area Roles | | | | $\Sigma$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Attacker | Neutral | Protagonist | Supporter | | Giver | Follower | Orienteer | Supporter | |
| | 0 | **.001** | .817 | .104 | .078 | 15k(1.0) | .162 | .777 | .043 | .018 | 15k(1.0) |
| # of spks | 1 | **.002** | .740 | .177 | .081 | 60k(1.0) | .251 | .675 | .049 | .025 | 60k(1.0) |
| | 2 | **.004** | .680 | .220 | .096 | 26k(1.0) | .325 | .591 | .054 | .031 | 26k(1.0) |
| | 3 | **.005** | .620 | .238 | .137 | 6k(1.0) | .358 | .536 | .070 | .037 | 6k(1.0) |
| | 4 | **.008** | .581 | .305 | .107 | 656(1.0) | .329 | .572 | .076 | .023 | 656(1.0) |
| $\Sigma$ | | 293 | 78k | 19k | 9k | 11k | 28k | 71k | 5k | 3k | 11k |

indicate that for 80% time in *Mission Survival Corpus I*, there are between one or two simultaneous speakers, an average of less than .88 protagonists (who drives the meeting) and between 1.004 to 1.3 information-givers. This is determined by the mental load of the participants and their conscious (or subconscious attempts) to increase the efficiency of the discussions. On the other hand, the fraction of the secondary roles, such as attackers, increases significantly.

We note that the statistical learning theory does not guarantee the learnability of features. Thus, we cannot treat a statistical learning method as a magical black box that takes training data as input and generates working models about the training data. What the theory provides instead is mathematically rigorous ways to avoid overfitting in statistical learning. As a result, our introspection into how we solve problems by ourselves and attain efficiency provides good intuitions about how our individual and collective mental processes can be "learned" and simulated by machines.

## 3.4.2　Individual Honest Signals

This section, introduces individual honest signals, their relationships to role-taking tendencies, their relevance with the roles, and their correlations.

### 3.4.2.1　Speech Features

Previous studies suggest speech can be very informative about social behavior. For instance, Pentland [Pentland, 2008] identified four classes of speech features for one-minute windows (Emphasis, Activity, Mimicry and Influence). He showed these classes are informative of social behavior and can be used to predict it. In Pentland's [Pentland, 2008] view, these four classes of features are honest signals, "behaviors that are sufficiently hard to fake that they can form the basis for a reliable channel of communication." To these four classes, Spectral Center may also be added, which has been reported to be related to dominance [Rienks and Heylen, 2005].

**Emphasis** is usually considered a signal of how strong the speaker's motivation is. In particular, its consistency is a signal of mental focus, while its variability indicates an openness to influence from other people. The features for determining emphasis consistency are related to the variations in spectral properties and prosody of speech: the less the variations, the higher the consistency. The relevant features are: (1) confidence in formant frequency, (2) spectral entropy, (3) number of

autocorrelation peaks, (4) time derivative of energy in frame, (5) entropy of speaking lengths, and (6) entropy of pause lengths.

The features for determining the **Spectral Center** are (1) formant frequency, (2) value of largest autocorrelation peak, and (2) location of largest autocorrelation peak.

**Activity** (=conversational activity level) is usually a good indicator of interest and engagement. The relevant features concern the voicing and speech patterns related to prosody: (1) energy in frame, (2) length of voiced segment, (3) length of speaking segment, (4) fraction of time speaking, (5) voicing rate (=number of voiced regions per second speaking).

**Mimicry** allows keeping track of multi-lateral interactions in speech patterns, and can be accounted for by measuring. It is measured through the number of short reciprocal speech segments, (such as the utterances of "OK?," "OK!," "done?" or "yup.").

Finally, **influence**, the amount of influence each person has upon another person in a social interaction, was measured by calculating the overlapping speech segments (a measure of dominance). It can also serve as an indicator of attention, since the maintenance of an appropriate conversational pattern requires attention.

For the analysis discussed below,windows of one minute in length were used. Earlier works [Pentland, 2008], suggested this sample size is large enough to compute the speech features in a reliable way, while being small enough to capture the transient nature of social behavior.

### 3.4.2.2 Body Gestures

Body gestures have been successfully used to predict social and task roles [Dong et al., 2007]. In this study, they were used as baselines to compare the import of speech features for socio and task roles prediction. Two visual features were consdered: (1) hand fidgeting and (2) body fidgeting. The fidgeting—the amount of energy in a person's body and hands—was automatically tracked using the MHI (Motion History Image) techniques. This methods uses skin region features and temporal motion to detect repetitive motions in the images and associate them with an energy value in such a way that the higher the value, the more pronounced is the motion [Chippendale, 2006]. These visual features were first extracted and tracked for each frame at a frequency of 3 hertz, then averaged out over the one-minute window.

### 3.4.2.3 Relationship between honest signals and role-taking tendencies

We note different people have different yet consistent styles in taking functional roles, and their styles are reflected in their honest signals.

The frequencies with which the 32 subjects assumed each of the eight functional roles in the first half of their respective discussions were compared with those in the second half, taking into account the fact it is difficult to collect a data set in which the same persons participant in many different types of discussions. The frequencies with which people take the Neutral/Follower roles, the Giver role, the Protagonist role, the Supporter role and the Seeker role in the first half of their discussions were predictive of the frequencies in the second half ($R^2 \geq .8, p < .001$). The frequencies in the first half were compared with those in the second half, but randomly permuted (within

discussion groups and within the whole data set). The permutations destroy both the correlations and the normality ($R^2 \leq .1, p \geq .3$). The consistency of taking the other roles in both halves of the discussions were weaker, but they nevertheless exist.

The hypothesis that the frequencies of taking different roles are linearly dependent on the honest signals with normal random noise was tested. The results show the rate a person takes the Seeker role is linearly dependent on his rate of short speaking segments ($R^2 = .47, p = .006$). The rate of taking the Supporter role is linearly dependent on the fraction of speaking and the fraction of voicing ($R^2 \geq .46, p \leq .007$)The rates of taking the Giver role, the Protagonist role and the Neutral/Follower roles are linearly dependent on most of the honest signals in the **activity** category.

## 3.5 Role Detection about *Mission Survival Corpus I*

Modern statistical methods normally guard against overfitting by several mechanism, and can normally attain comparable performances by careful selection of features and careful formulation of problems. However, some methods may be easier to use and more intuitively understandable for some problems. This Section discusses the one-person features and the interaction features that statistical learning methods (the support vector method and the influence model, in particular) should utilize in order to obatin good performance, the different ways the methods use the features, and the resulting performances.

The turn-taking behavior related to the functional roles of a speaker at a specific moment include: (1) his amounts of speaking time in time-windows of different sizes around the moment; (2) whether other persons oreinted their bodies to the speaker at the beginning/end of his speaking turn; (3) the amounts of speaking time of other persons in time windows of different sizes around the moment, specifically, the amounts of speaking time of the persons who speak the most in those time windows; and (4) the psychological profile of this person, e.g., his extrovertedness, his tendency to take control and his level of interest in the discussion topic.

The influence model formulates the group process in terms of how an individual takes his functional role based on the functional roles of the others on the one hand, and how an individual presumes the others' possible functional roles based on everyone's current roles on the other hand (c.f. Figure 3.2). When a person is taking the Giver role, he prefers others to take the Neutral/Follower role or the seeker role at least for a while. In comparison, when a person takes the Neutral role, he does not usually care who is going to take which role next. When all participants take the Neutral role, the overall preference of the whole group can be very weak. When this happens, the individuals can wander about their role-taking states until some individual takes a "stronger" role. Specifically, an influence model can determine which functional role a participant is most likely to take at a given moment by comparing how likely different roles correspond to his amounts of speaking time in time-windows of different sizes around this moment. For example, when there are doubts about whether a person is shifting to the Giver role or the Seeker role, the influence model will examine the intensity of the other participants' body movements. The Giver role is associated with more body movements at the beginning of the corresponding speaking-turn and more attention from the others. The role which had been taken by a participant a moment previously can be used by the influence model to generate a "vote" for different roles for the participant
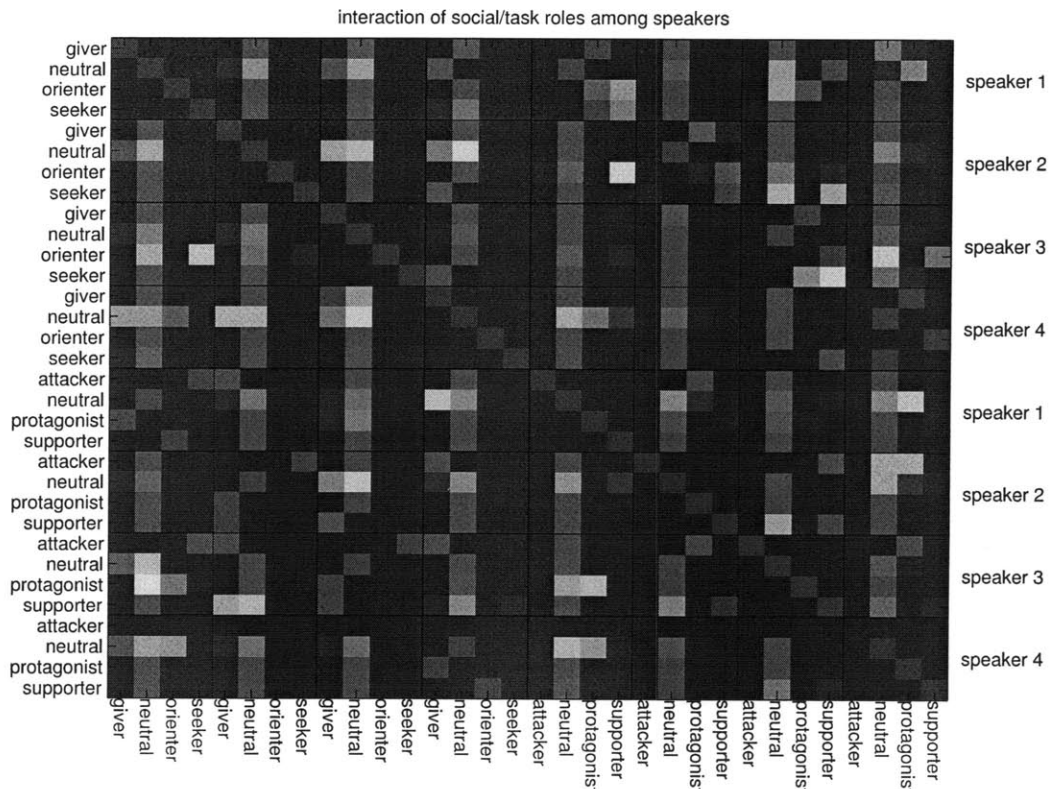
Figure 3.2: Roles combine with each other in useful ways,hence we can infer roles from the roles of others.

under investigation This vote can subsequently be used to bias the model's Bayesian estimation. The psychological profile of the participants can be further used for generating the votes (for the participants to take certain roles).

In contrast, the support vector method (SVM) does not involve a probability distribution in the training phase and the application phase, although SVM can be used within the Bayesian frame-work. SVM also requires the model observations to be points in a (possibly high-dimensional) Euclidean space. In terms of utilizing the amounts of speaking time corresponding to different window sizes, the SVM performs as well as any Bayesian method, and the latter requires appropriate probability estimations of the observations conditioned on the functional roles. The amounts of speaking time were sorted over all participants in every time-window of different sizes up to some upper bound (two minutes in these experiments).The sorted amounts of speaking time over all speakers and corresponding to all window sizes around the moment of inspection (among other features) were used for functional-role classification. The arrangement by sorting makes the corresponding feature permutation independent. The SVM can subsequently disambiguate among the possible roles of a person by comparing his amounts speaking time with those of the others, particularly by comparing them with those of the person who speaks the most. The hand/body movements involved with role-shifting is the most difficult feature for the sSVM to deal with, since the boundaries of the role assignments are unknown. Functional-role classifiers who have

Table 3.4: Performances of classifying task/social-emotional roles using SVM/IM with interaction signals.

| | | SVM | | | | | influence model | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Giver | Follower | Orienteer | Seeker | % | Giver | Follower | Orienteer | Seeker | % |
| ground truth | G | **10758** | 1944 | 982 | 1092 | .72 | **10173** | 1441 | 880 | 2282 | .68 |
| | F | 2581 | **26924** | 334 | 3401 | .80 | 1542 | **28000** | 246 | 3542 | .84 |
| | O | 796 | 153 | **795** | 447 | .36 | 708 | 260 | **1045** | 178 | .48 |
| | S | 362 | 639 | 27 | **277** | .21 | 426 | 219 | 32 | **628** | .48 |
| | % | .74 | .90 | .37 | .05 | | .79 | .94 | .47 | .10 | |

| | | SVM | | | | | influence model | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Attacker | Neutral | Protagonist | Supporter | % | Attacker | Neutral | Protagonist | Supporter | % |
| ground truth | A | **95** | 63 | 15 | 12 | .51 | **108** | 54 | 5 | 18 | .58 |
| | N | 452 | **32779** | 3932 | 1308 | .85 | 450 | **32813** | 3916 | 1292 | .85 |
| | P | 320 | 813 | **5789** | 2458 | .62 | 318 | 804 | **5804** | 2454 | .62 |
| | S | 143 | 251 | 1738 | **1344** | .39 | 142 | 258 | 1733 | **1343** | .39 |
| | % | .09 | .97 | .50 | .26 | | .11 | .97 | .51 | .26 | |

been trained in the use of SVM indicate that SVM uses the body/hands movement corresponding to the current moment in order tor disambiguate the roles.

Table 3.4 compares the performance of the influence model and the performance of SVM using the best construction of features were capable of generating. The performances of both SVM and the influence model are seen to be improved compared with a my previous study[Dong et al., 2007], especially with regards to the infrequently-appearing roles. The improvement is due to a better understanding of the group process. In this present study, the influence model displayed a slightly higher performance than the SVM, in contrast to the previous study in which the influence model performed somewhat more poorly than the SVM. This is due to the new perspective and corresponding EM algorithm for the influence modeling. In this current study, the latent state space of an influence model was the summed voting of the individuals' future states (e.g., participants' next functional roles) associated with a probability space, whereas in the previous study, it was the marginal probability distributions among those states. A direct consequence of this new perspective is that, previously, a person taking a neutral role could not waive his votes, in the current study, they were able to do so. .

# 3.6 Group Process Statistics and Performances in *Mission Survival Corpus I*

One reason to investigate group processes and the Task/Socio-Emotional Area roles is to facilitate the design of automated tools to improve the performance of the group. In the Mission Survival Corpus I, the initial individual scores and the final group scores of seven discussions out of a total of eight are available. and they are in terms of how the individual/group rankings of 15 items

are different from the standard expert ranking. ($f(r_1 \cdots r_{15}) = \sum_{i=1}^{15} |r_i - r_i^{(0)}|$, where $f$ is the score function, $r_1 \cdots r_{15}$ is an individual/group ranking, $r_1^{(0)} \cdots r_{15}^{(0)}$ is the expert ranking, $r_i \neq r_j$ and $r_i^{(0)} \neq r_j^{(0)}$ for $i \neq j$, and $r_i, r_i^{(0)} \in \{1, \cdots, 15\}$.) Thus, the corpus provides laboratory data with which it is possible to examine how individuals with different initial performances and psychological profiles interact with each other. This makes it possible for them to incorporate their individual information to attain better performance. The preliminary findings are given below.

The post-discussion group performance is linearly and positively correlated with the average of the pre-discussion individual performances of the participants, with the pre-discussion performance being slightly better than the post-discussion performance. (Group score = .93× average of individual scores -.74, $R^2 = .58$, $p = .03$). The relationship is shown in Figure 3.3 (a), and can be explained with a probabilistic model of how individuals combine their results. Prior to the discussion, pieces of information for solving the ranking problem of the mission survival task are probabilistically distributed among the participants, and different individuals have the correct/best rankings for different items. During the discussion, the individuals merge their information through a group process that is probabilistically dependent on their initial performances, their interactions with each other, and numerous other factors. When the individuals disagree on the ranking of an item, they can either choose one from their repository of rankings that results in minimal disagreements, or find a creative and better ranking by further sharing of information and an "aha" experience. Previous experiments have shownthat groups make use of their resources to a probabilistically similar extent, and experienced discussion groups outperform inexperienced groups by generating more item rankings that are more creative and more correct as as a result of better information sharing [Hall and Watson, 1970]. Thus, the relationship between the group performance and the average of the individual performances follows from the fact the different groups are more or less similar.

The improvement of a group's performance over the average performance of the individuals within is positively correlated with the amount of speaker-overlap throughout the discussion, which reflects the intensity of the discussion ($R^2 = .35$, $p = .10$). The relationship is shown in Figure 3.3 (b). The improvement is also positively correlated with the frequency with which meeting participants take the protagonist role ($R^2 = .35$, $p = .10$) as well as with the rate at which meeting participants take the giver role ($R^2 = .28$, $p = .12$). Since the average length of a continuous protagonist role segment is almost 50% longer than the average length of a continuous giver role segment (26 vs. 19), I speculate that the improvement is more dependent upon the longer utterances of the individuals. The improvement doesn't seem to correlate with either the length of a discussion or the rates at which the participants take other roles. These correlations are again compatible with the observation reported in previous experiments that an experienced discussion group encourages better problem solving and more thorough information sharing [Hall and Watson, 1970]. An inexperienced discussion group worries more about whether it can eventually reach a consensus. As a result, its members treat consensus-reaching as the goal, rather than the natural result of sufficient information sharing. They either argue for their own rankings without paying attention to the arguments of other group members , or give up their rankings too easily. In either case, they feel their importance in the discussion is not being sufficiently recognized, and quickly lose their motivation for participation. In contrast, an experienced group encourages different opinions, and views conflicts as evidence of insufficient information sharing. Its participants solve a conflict by a thorough discussion and a win-win problem solving strategy rather than by superficial or me-

grp perf. ~ avg. perf ($R^2$=.58,p=.03)          perf. impr. ~ spkr overlap ( $R^2$=.35,p=.10)
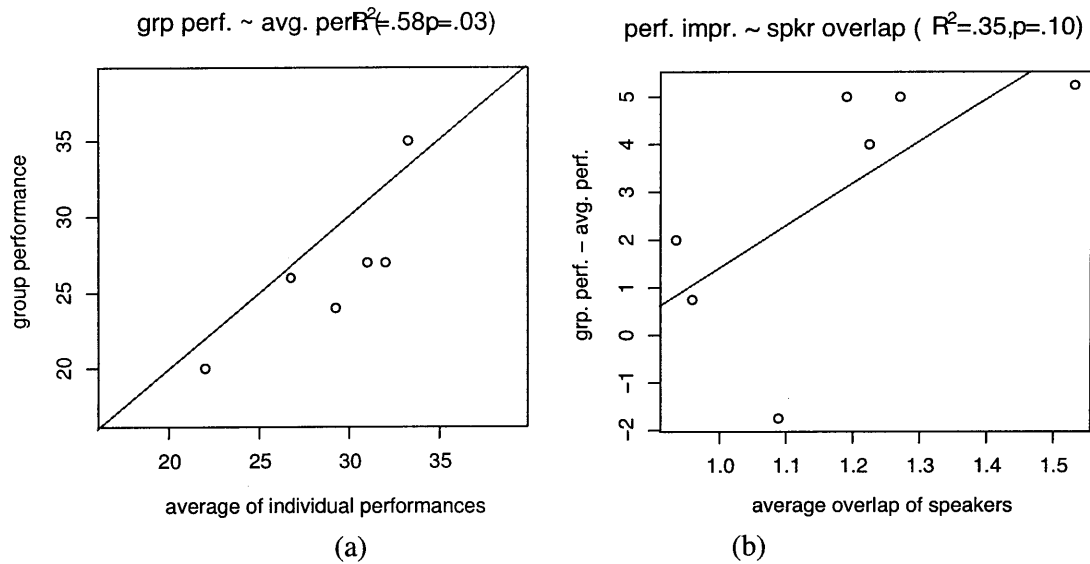
Figure 3.3: (a) Group score is linearly related with the average of pre-discussion individual scores. (b) Group performance is linearly related with the average overlap between speakers. This can be seen as a measure of engagement in the task or the intensity of the discussion.

chanical techniques such as majority-rule voting and coin-flipping. The participants also display an appropriate degree of skepticism and suspicion about initial agreements that seem to be arrived at too easily. Different group discussion process characteristics and different role-playing behavior can be dependent upon the different opinions between an experienced group and an inexperienced group about what constitutes fruitful group discussion.

In many of the discussions, one or a few individuals take certain tasks and social roles twice as frequently as they do other roles. The fraction of speaking time of an individual and the rates that an individual takes the giver and/or the protagonist role do not seem to correlate with the initial performance of the individual. Based on the small number of discussions and the fact that the meeting corpus is in Italian, we can only speculate that role-taking is related to the personal styles of the individuals, their motivation, their interaction with each other, as well as numerous other factors.

While it is my belief that the symptoms of group process problems could be found by automated tools and the prescriptions could be given accordingly, I note that facilitating the group process is considerably more difficult. One reason for this is that an inappropriately-phrased prescription may inadvertently distract the attention of the participant from the real goal of the discussion, and shift their attention from one unimportant concern (e.g., the pressure of reaching a consensus) to another unimportant concern (e.g., the "appropriate" speaking-turn lengths and the "appropriate" amount of speaker overlap).

## 3.7 Summary and Discussions about the Mission Survival Experiment

This chapter discussed the turn-taking dynamics and the changing individual role assignments of several group brainstorming sessions on different open problems. It also discussed their modeling and learnability issues using several statistical learning methods (support vector machines, the hidden Markov model, and the influence model). The group discussion dynamics were first modelled by first thinking about how such discussions should work to acheive their purposes, and then by applying the appropriate statistical learning methods. There are several possible future directions:: First, it may prove interesting to simulate the behavior and performances of different types of brainstorming sessions with stochastic processes and simplified assumptions, then compare the simulated results with the experimentally collected results. The simulation could provide insight for understanding collective intelligence. Secondly, I would like to use modeling to improve the efficiency of multi-person interaction.. Thirdly, I would like to know whether such turn-taking and role-assignment modeling could be suitable for other types of multi-person interactions with appropriately tuned parameters.

# Chapter 4

## Improving Group Performance with Meeting Mediators

Modeling and facilitating group problem-solving are key issues in many fields including management, cognitive sciences, human computer interaction and robotics. I will discuss our novel approach of mapping the interaction patterns of group processes to their performances. We use sensors to automatically measure group behavior and actions. We also see how the modification of group dynamics changes the resulting performance. Our discussions are based on a lab study data set using the meeting mediator system through which we collected objective quantitative data. I believe the findings of this data set are applicable to many real world task-group processes.

While it is an important topic in cognitive sciences, the study of the relationship between interaction and performance is often suffered from its non-analytical nature and is often marginalized to some extent. I will base the discussion on the **meeting mediator data set** [Kim et al., 2008], so that we can reason the group problem-solving strategies and quantitatively analyze the relationship between the group performances and the group dynamics. The group brainstorming and decision making processes in the meeting mediator data set are representative of the group processes in many task groups.

The **meeting mediator** is a system that detects and displays group dynamics. Its goal is to quantitatively measure the group's interaction pattern and provide real-time feedback to promote change in the group's behavior. The meeting mediator system consists of two components: the **sociometric badge** and a mobile phone. The sociometric badge is an business-card-sized embedded device, that understands the behavior and interaction of the sociometric badge user by collecting the audio and motion information of him and interact with other sociometric badges through radio, IR and BlueTooth channels. It is the input and the brain of a meeting mediator. The phone displays group interactivity level and participation balance which is transferred from the badges via Bluetooth connection. A more detailed description and pictures of a meeting mediator can be found in the work of Kim et al. [Kim et al., 2008].

The meeting mediator data set involves over 40 group problem-solving processes in different configurations. Our goal in collecting this data set is to understand the relationship between the task group dynamics and the task group performances, so that we can design automated tools to estimate and improve group performances using signals that can be reliably detected and estimated with embedded-device hardware/firmware.

The two tasks in the group processes are two 20-questions games [20q]. We split each task into two parts: Each group of four persons is first required to brainstorm as many ideas as possible that is compatible with a partially finished game — a list of 10 yes/no questions and the corresponding answers — in eight minutes (c.f. Table 4.1). The group is then required to ask as few questions as possible through its leader to get the answer. In order to solve a 20-questions game by asking the fewest number of questions, the group has to construct each question carefully based on its estimation of the items that satisfy the answers to all previously asked questions. As a result, using the two-parts form for the tasks both enforced a good structure in the group problem-solving processes and gives us extra semantic clues on why some groups asked fewer questions than others to accomplish the tasks.

The data about each group process in the meeting mediator data set consist of: (1) the features about the group dynamics and captured by the **sociometric badges** [Olguin et al., 2009] worn by the experiment subjects, in particular at each sample time whether an individual subject is speaking, how he moves his body, and how he orients his body in relation to the others; (2) the facts related to the group performances, including the ideas generated by each team/individual in the brainstorming sessions, the times spent for every single groups to generate every single questions; and (3) the surveys, in particular those about the subjects' own opinions on the group dynamics and group performances. We neither audio-recorded nor video-recorded any of the group processes, following the COUHES guidelines [COU] to protect the privacy of the experiment subjects.

Each group was instructed that collaboration was important to group performances prior to its tasks. The task-groups were randomly assigned one of four labels: A, B, C and D. To test the effect of feedback on the groups' behavior, the members in the groups labeled with C or D were not allowed to see the feedback of the meeting mediators. In the following, we will simply refer the groups labeled with C or D as groups without meeting mediators.

## 4.1   20-Questions Game

Facilitating group problem solving in 20-questions games is possible because the 20-questions game is amiable for group problem-solving, the performance of a task-group in solving this game could be reliably estimated from the group turn-taking dynamics, and a meeting mediator could direct its user to a high level of participation based on its understanding of the group process. Before we inspect the groups in the meeting mediator data set and reason how they could improve their performances, let us make a plan on how we would like to solve a 20-questions game with the best possible performance by ourselves.

In a 20-questions game, the answerer has a thing in his mind, and the questioner is responsible for pinpointing the thing with as few yes/no questions as possible. By information theory, a good question should eliminate half of the remaining candidates that are compatible with the previous question/answer pairs. Since the questioner does not know the set of all compatible candidates a priori, he has to choose a good question based on his sampling of the set. Thus the more similarly the answerer and the questioner sample the candidate sets, the better performance the questioner will have.

We can also put the previous paragraph in the language of probability: The answerer samples a thing with a probability measure $\mathbb{P}$; The questioner reduces half of probabilities at each step assuming he can, or $\mathbb{Q}(S_{t+1}) = \frac{1}{2}\mathbb{Q}(S_t)$ where $S_t$ represents the set of all compatible candidates at step $t$ and $\mathbb{Q}$ is the probability measure used by the questioner; The closer $\mathbb{P}$ is to $\mathbb{Q}$, the better performance the questioner will have; The questioner samples the set of compatible candidates according to a diffusion process implicit in his mind $u_t^{(i+1)} \sim \mathbb{Q}(\bullet | \{u_{t'<t}^{(j)}\} \bigcup \{u_t^{(1)}, \cdots, u_t^{(i)}\}) \uparrow S_t$ and ask questions based on his samples $S_{t+1} = S_t \bigcap \chi(\{u_{t'<t}^{(j)}\} \bigcup \{u_t^{(1)}, \cdots, u_t^{(m_t)}\})$ where $u_t^{(i)} \in S_t$ is an item that the questioner sampled at time $t$, and the characteristic function $\chi$ is what the questioner subsequently asked. Solving a 20-questions game with a group of people provides wider diffusion, larger sample sizes at each steps, and better questions. When the questioner do not know the $\mathbb{P}$ in the mind of the answerer, his safest bet is the uniform distribution over the set of all things, and he is expected to pinpoint the thing in $\log |S|$ steps.

I will give the mechanism of brainstorming and the formula to predict decision making performance from the corresponding brainstorming performance, both referring the meeting mediator data set, show how the group dynamics collected by the meeting mediators determine the brainstorming performance and consequently the decision making performance, and show how the meeting mediators could improve group performance.

# 4.2 Brainstorming & Decision Making

In this section, we will discuss how the groups in the meeting mediator data set came to new ideas in the brainstorming sessions, and how the groups subsequently made decisions based on their brainstorming results.

## 4.2.1 Brainstorming

A brainstorming session such as one in the meeting mediator data set can be viewed as a diffusion process. The subjects in the session randomly hit new ideas based on the ideas they have recollected due to familiarity or accessibility and the ideas that they have already considered. The ideas are then checked against the given constraints — a partially finished game represented by a list of yes/no questions together the corresponding answers. If the group hits one idea, it will be more likely to hit other closely related ideas. From this perspective, the set of all ideas that are compatible with the given constraints are related with each other according to the accessibility from one idea directly to another idea, and the inspected ideas of a group diffuse on the network of all ideas. Based on the diffusion argument, a group of people working together will statistically generate more ideas than the same people working individually assuming all other conditions are the same, since in the former configuration people can inspire each other and thus the edges connecting different items are more accessible. Based on the same argument, a group of interacting strangers will statistically hit more ideas than a group of interacting friends, since in the former configuration the different backgrounds of the strangers imply more "unexpected" edges from items to items [Wilson et al., 2004].

With the knowledge of which group hit which ideas (c.f. Table 4.2 and Table 4.3), the structure of the space of brainstorming ideas can be computed: As we have discussed previously, a group either generated all/most elements in a set of closely related items or none/few in this set. In other words, there is a equivalence relation on the set of all items for this group. Thus the set of items generated by a group is the union of some sets of statistically related items (which forms equivalent classes) from the perspective of the group, so is the set of items not generated by the group. When we change our perspective and inspect how the items were generated by the groups, rather than how the groups generated the items, a distance among the items can be defined: Two ideas share an obvious relationship and are thus close to each other if most of the groups generated either both or none of them; Two ideas are related by uncommon relationships and are thus far from each other if many groups generated only one of them. The space of ideas, represented by the set of ideas and the set of relationships among the ideas, can subsequently be mapped into a familiar structure, such as a Euclidean space, a dendrogram or a graph, and be inspected.

In the meeting mediator data set, 52 groups have their task 2 brainstorming outcomes recorded, and they generated 92 distinct items in total. The overall result can be represented as a $92 \times 52$ matrix, with the matrix rows corresponding to distinct items and the matrix columns corresponding to different teams. The entry that lies in the i-*th* row and in the j-*th* column is either 0 or 1, representing group $j$ did or did not generate idea $i$ in its brainstorming.

The two characteristics of the dendrogram in Fig. 4.1 are listed below. They have implications on how a 20-questions game is played, and how the game-play can be modeled, as will be discussed later. First, the brainstorming ideas of task 2 in the meeting mediator data set form small clusters. Of these clusters, some consist of ideas that truly go together ("stapler" and "hole puncher"); and a few — "clusters invented by a clever group" — consist of ideas that were hit by a single group. Second, the small clusters are well separated from each other. As a result, the idea set of a partially finished game can be split into a yes-set and a no-set equally well in many ways, assuming the corresponding yes/no questions can be found. The well-separateness of the small clusters also means that which cluster of ideas a group is going to hit next is more or less unpredictable.

If we define an Euclidean (i.e., a 2-norm) distance between the brainstorming ideas and subsequently do eigenvalue decomposition on the resulting distance matrix, we will find that the largest eigenvalue accounts for more than 80% of variance, the corresponding eigenvector sorts the ideas based on the frequencies they were hit, and the rest eigenvalues are relatively equivalent to each other. Thus the observation from principle component analysis coincides with the second characteristic we just described based on the dendrogram.

Known the structure of the brainstorming ideas, we could be able to inspect how a group hit new ideas as a function of time, as well as how a group of interacting strangers take unexpected paths and statistically hit ideas faster than a group of interacting friends or a group of people working alone.

## 4.2.2 Decision Making

The brainstorming session of a task gives a group a considerable amount of time (8 minutes out of 8+10 minutes) in sampling the set of candidate items. It also implicitly prepares the group on

Table 4.1: Task #1 Clues & Task #2 Clues.

- Task #1 Clues: Thing

    1. 1. Is it shiny?.................................................................................................No
    2. 2. Do you hold it when you use it?.................................................................Yes
    3. 3. Can it fit in a shoebox?............................................................................Yes
    4. 4. Would you use it daily?............................................................................Yes
    5. 5. Is it flexible?...........................................................................................Yes
    6. 6. Is it decorative?.......................................................................................No
    7. 7. Does it open?..........................................................................................No
    8. 8. Is it found in the home?...........................................................................Yes
    9. 9. Do you clean it regularly?........................................................................Yes
    10. 10. Is it organic?........................................................................................No

- Task #2 Clues: Thing

    1. Is it used for entertainment?.......................................................................No
    2. Would you give it as a gift?.........................................................................No
    3. Is it a tool?.................................................................................................Yes
    4. Is it smaller than a shoebox?......................................................................Yes
    5. Would you find it in a toolbox?...................................................................No
    6. Does it use electricity?...............................................................................No
    7. Is it assembled?.........................................................................................Yes
    8. Is it meant to get wet?................................................................................No
    9. Does it have metal parts?............................................................................Yes
    10. Is it used indoors?.....................................................................................Yes

Figure 4.1: The dendrogram of the items generated in task 1 brainstorming and task 2 brainstorming. If two items go together, both or neither of them were generated by most groups.

Table 4.2: The brainstorming outcomes (task 1) of different groups reveal the regularity and diversity human thinking process.

| | A.01 | A.02 | A.03 | A.04 | A.05 | A.06 | A.07 | A.08 | A.09 | A.10 | A.11 | A.12 | A.13 | B.01 | B.02 | B.03 | B.04 | B.05 | B.06 | B.07 | B.08 | B.09 | B.10 | B.11 | B.12 | B.13 | C.01 | C.02 | C.03 | C.04 | C.05 | C.06 | C.07 | C.08 | C.09 | C.10 | C.11 | C.12 | C.13 | D.01 | D.02 | D.03 | D.04 | D.05 | D.06 | D.07 | D.08 | D.09 | D.10 | D.11 | D.12 | D.13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Brillo Pad | | | | | | | | | 1 | | | | | | 1 | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | 1 | | | | | | | | | |
| Brush | 1 | | 1 | 1 | | 1 | | 1 | 1 | 1 | | | | | | | | 1 | 1 | 1 | 1 | | | 1 | 1 | 1 | | 1 | | | | | | | | | 1 | | | 1 | | 1 | | | | | | | | | | |
| Comb | 1 | 1 | 1 | 1 | 1 | 1 | | | | 1 | | | 1 | 1 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | 1 | 1 | 1 | | 1 | | 1 | 1 | | | | | | | | | 1 | | 1 | | | | | | | | | 1 | 1 |
| Contacts | | 1 | | | | | | | | 1 | | | | | | | | | | | | | 1 | | 1 | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | |
| Cup | | | 1 | | 1 | 1 | 1 | | | | | | | | 1 | 1 | | | | 1 | | | 1 | | | | | 1 | | | 1 | | 1 | | 1 | | | | | | | | | | | | | 1 | | | | |
| Cutting Board | | | | | | | | | | | | | | 1 | | | | | | | | | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | |
| D...o | | | | | | | | | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Dish Rag / Towel | | | 1 | | | 1 | | 1 | | 1 | | | | | | | | 1 | 1 | | | | | 1 | | | | | 1 | 1 | 1 | | | | | 1 | | | 1 | | 1 | 1 | 1 | | 1 | | 1 | | | | | |
| Dishwash Glove | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Eraser | | | 1 | | | | | 1 | | 1 | 1 | 1 | | 1 | | | | | | | | | | | | | | | | | | | | | | 1 | | | | 1 | | | 1 | | | | 1 | | | | | |
| Facial Cloth | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Feather Duster | | 1 | 1 | | | | | | | | | | | | | 1 | | 1 | | | | | 1 | | | | | | | | | | | | | | | | | 1 | | | | | | | | | 1 | | | |
| Flexible Muffin Tin | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Flexible Pan | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Glove | | 1 | 1 | | 1 | | 1 | | | | | | | | | | | 1 | | 1 | | | | | | | 1 | | | | | | | | | | | | | 1 | | | 1 | 1 | | 1 | | | | | |
| Hair Band | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Hairbrush | 1 | 1 | | | 1 | 1 | 1 | 1 | | | | | | 1 | 1 | 1 | 1 | 1 | | | 1 | | 1 | | | | 1 | 1 | 1 | | 1 | 1 | | | 1 | | 1 | | 1 | 1 | | 1 | 1 | 1 | | 1 | | 1 | 1 | |
| Hand Broom | | | | | | | | | | | | | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Hand Mop | | | | | | | | | | | | | | 1 | | | | 1 | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Handkerchief | | | | 1 | | 1 | | | | | | | | | | | 1 | | | | | | | | | | | | | | | | | 1 | | | | | | | 1 | 1 | | 1 | | | | | | | | |
| Lidless Tupperware | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | | | | | | | | | | | | | | | | | | |
| Lint Brush | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | | | | | | | | | | | | | | | | | | |
| Lint Remover | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Loofah | | 1 | | | | 1 | | 1 | | | | | | | | 1 | | | | | | | | 1 | 1 | | 1 | | | | | | | | | | | | | 1 | 1 | | | | | | | | | | | |
| Napkin | | | 1 | | 1 | | | | | | | | | 1 | | | | | | 1 | | | | | 1 | 1 | | | | | | | | | | | | | | 1 | | | | | | | | | | | | |
| Oven Gloves | | 1 | | | | 1 | 1 | | | | | | | | | | | | | | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | |
| Place Mat | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Plastic Dishware | | 1 | | | | | | | | 1 | | | | | | | | | | 1 | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | | | |
| Pot Holder | | 1 | 1 | | | | | | | | | | | | | | | | | | | | | | | | 1 | | 1 | | | | | | | | | | | 1 | | | | | | | | | 1 | | | |
| Shoe Brush | | | | | | | | | | | | | | 1 | | | 1 | | | 1 | | | 1 | | | | | | | | | | | | | | | 1 | | | | | | | | | | | | | | |
| Shower Cap | | 1 | 1 | | | | | | | | | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Shower Curtain | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | | | | | | | | | | | | | | | | |
| Shower Pouff | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | | | |
| Soap Brush | | | | | | | | | | | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Spatula | 1 | | 1 | 1 | | | 1 | | | | | | | | | | | | 1 | 1 | | 1 | | | | 1 | 1 | | 1 | | 1 | | 1 | | | | | 1 | | 1 | | | | 1 | 1 | | | | | | | |
| Sponge | 1 | | 1 | | 1 | 1 | 1 | 1 | 1 | | | | | 1 | 1 | 1 | | 1 | 1 | 1 | 1 | | 1 | 1 | 1 | 1 | | 1 | | 1 | 1 | 1 | 1 | | 1 | | 1 | 1 | | 1 | 1 | 1 | 1 | 1 | 1 | | 1 | | | | | |
| Stress Ball | | 1 | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | | | |
| Swiffer | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | 1 | | | | | | | | | 1 | | | | | | | | | | | | | | | | |
| Toilet Brush | | | | | | 1 | | 1 | | | | | | | | | | | | 1 | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Toothbrush | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | 1 | 1 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | 1 | 1 | 1 | | 1 | 1 | 1 | 1 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | |
| Towel | 1 | 1 | | 1 | 1 | 1 | 1 | 1 | 1 | | 1 | | 1 | 1 | 1 | 1 | 1 | | 1 | 1 | 1 | | 1 | 1 | 1 | | 1 | | 1 | | 1 | | 1 | 1 | 1 | 1 | 1 | | 1 | 1 | 1 | 1 | 1 | 1 | | | | | | | | |
| Tupperware Box | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Tupperware Lid | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Washcloth | 1 | 1 | | | 1 | | | 1 | | | | | | 1 | 1 | 1 | 1 | 1 | 1 | | 1 | 1 | 1 | | | | 1 | 1 | 1 | | 1 | 1 | 1 | 1 | 1 | 1 | | 1 | | 1 | | 1 | | 1 | | 1 | 1 | | 1 | | | |
| Watch | | 1 | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | | | | | | | | | | | | | | | | | |
| Whisk | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | | | | | | | | | | | | | | | | | | |

| | A.01 | A.02 | A.03 | A.04 | A.05 | A.06 | A.07 | A.08 | A.09 | A.10 | A.11 | A.12 | A.13 | B.01 | B.02 | B.03 | B.04 | B.05 | B.06 | B.07 | B.08 | B.09 | B.10 | B.11 | B.12 | B.13 | C.01 | C.02 | C.03 | C.04 | C.05 | C.06 | C.07 | C.08 | C.09 | C.10 | C.11 | C.12 | C.13 | D.01 | D.02 | D.03 | D.04 | D.05 | D.06 | D.07 | D.08 | D.09 | D.10 | D.11 | D.12 | D.13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Abacus | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Alarm Clock | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Apple Peeler | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | |
| Bagel Slicer | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | |
| Balance | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | | | | | | | | | | |
| Belt | | | | | | | | 1 | | | | | | | | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | 1 | | | | |
| Bike Pump | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | |
| Binder | | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Binding Clip | | | | | | | | | | | | | | | | | | | 1 | | | | | | 1 | 1 | | | | | | | | | | | | | | 1 | | | | | | | | | | | | |
| Book Ends | | | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Can Opener | | 1 | | 1 | 1 1 1 1 | | | | | 1 | | 1 | | | 1 | | 1 | 1 1 | | | | | | | | | 1 1 1 | | | | 1 | | | | 1 1 | 1 1 1 | | | | | | | | 1 1 1 | | | | |
| Candle Holder | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | |
| Carabiner | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | 1 | | | | | | | | | | | | |
| Cheese Grater | | | | | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | 1 | | | | | | | | 1 | | | | 1 | | | | | | | |
| Clock | 1 | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | 1 | | | | | 1 | | 1 | | 1 | | | | | | | | | | |
| Compass | | | | | | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Corkscrew | | 1 | | 1 1 | | | | 1 | | 1 | | 1 | | 1 | 1 | 1 | 1 1 1 | | 1 | | | 1 | | | | | | 1 | | | | 1 | | 1 | | 1 | | 1 | | | | | | | 1 1 | | | | |
| Corn Holder | | | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Cosmetic Box | | | | | | | | | | | | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Coupon Holder | | | | | | | | | | | | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Desk Chair Wheel | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Door Hinge | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Door Knob | 1 1 1 1 | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | |
| Ear Piercer | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | | | | |
| Egg Beater | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | | | | |
| Egg Slicer | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | | | | | | | | | | | |
| Egg Timer | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | | | | | | | | | | | |
| Eyelash Curler | | | | | | 1 | | 1 | | | | | | | | | | | | | | | | | 1 | | | | | | | | | | | | | | 1 | | | | | | | | | | | | |
| Fire Extenguisher | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Fountain Pen | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Garlic Crusher | 1 | | | | | | | | 1 1 | | | 1 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Glasses | 1 | | | 1 | | | 1 1 1 | | 1 | | | | | | | | | 1 | | | | | | | | 1 | | | | | | | | | | | | 1 1 | | | | | | | | | | |
| Gun | 1 | | | 1 | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Hand Crank Radio | | | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Hand Grinder | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Hand Vac | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Hem Ripper | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Hole Punch | 1 | | | 1 1 1 1 | | | | | 1 1 1 | | | | | 1 | | | 1 | | 1 1 | | | | | 1 | | | 1 | | 1 | | | 1 | | | 1 | | | | | | | | | | | 1 | | | | |
| Hourglass | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Ice Cream Scooper | 1 | | | | | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Iron | | | | | | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Juicer | | | | | | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Kazoo | 1 | | | | | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Key Holder | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | |
| Kitchen Tongs | 1 1 | | | | | | | | 1 | | 1 | | 1 | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Labeller | | | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Leash | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | | | | |
| Lemon Zester | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | | | |
| Letter Opener | | | | | | | | 1 | | | | 1 | | | | | | | | | | | | | 1 | | | | 1 | | 1 | | | 1 | | | 1 | | 1 | | | | | | | | | | | | |
| Lighter | 1 | | | 1 | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Lobster Cracker | | 1 | | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Magnifying Glass | 1 | | | 1 | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | | | | | | | | | | | |
| Mech. Pencil | 1 | | 1 1 | | | | | | | | | | 1 | | | | | | | | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | |
| Metal Stamper | | | | | | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | | | | |
| Metronome | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | | | | | |
| Microscope | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Mousetrap | 1 1 | | | | | | | | | | | | | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Nail Clippers | | | 1 1 1 | 1 | | | | | | | 1 1 | | 1 | | | | | | | | | | | | | | | | | 1 | | | | | | | 1 | | 1 1 1 | | | | | | | |
| Nail File | | | | | | | | | | | 1 | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | | | | | |
| Nail Gun | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Nose Hair Trimmer | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | | | | | | | | 1 | | | | | | | | | | | |
| Nut Cracker | | 1 | | | 1 | | | | | 1 | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | 1 | | | | 1 | | | | | 1 | | | | | |
| Parer | | | | | | 1 | | | | | | | | | | | 1 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Padlock | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Pen | 1 1 1 | | | 1 1 1 1 | | | | | 1 | | 1 1 | | 1 | | | | | | 1 1 1 1 | | | | | 1 1 | | 1 | | | 1 | | | | | 1 1 1 | | | | |
| Pen Holder | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Pencil Sharpener | | | | | | | | | | 1 1 | | 1 | | | | | | | | | | | | | | | 1 1 1 1 | | | | | 1 | | | | 1 | | | | | | | | |
| Pepper Grinder | | | 1 | | | 1 | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Pizza Cutter | | | | | | | | | | | 1 | | | | | | | | 1 | | | | | | | | | | 1 1 | | | | | | | | | | | | | | | | | | | | | | |
| Pocket Knife | 1 | | | 1 | | 1 | | | 1 1 1 | | 1 | | 1 1 1 | | 1 | | | | 1 | | | | | | 1 1 | | 1 | | | 1 1 | | | | |
| Rolling Pin | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Safety Pin | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | | | |
| Salt Shaker | | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | | | | | | | | | | | |
| Scale | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Scissors | 1 1 | | | 1 1 1 1 1 | | | | | 1 1 1 1 | | | | | 1 1 1 1 1 | | 1 | | 1 1 | | | 1 1 1 1 | | | | 1 1 1 | | 1 1 1 1 1 | | |
| Shears | | | | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Shoe Horn | | | | | | | | | | | | | | | | 1 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Slide Rule | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | |
| Spiral Notepad | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | |
| Spray Paint Can | | | | | | | | | | | | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Staple Gun | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Staple Remover | 1 | | | | | | | | | | | | | | | | 1 | | | | | | | | | | 1 | | | | | 1 1 | | | | | | | | 1 | | | | | | | | | |
| Stapler | 1 1 | | | 1 1 1 1 1 | | | | | 1 1 1 | | 1 | | 1 1 | | | 1 | | | | 1 1 1 1 1 | | 1 | | | 1 | | | | | | | |
| Stethoscope | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Syringe | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Tape Dispenser | | | | 1 | | | | | | 1 | | | 1 | | | | | | | | | | | | | | 1 1 | | | | | | | | | | | 1 | | | | | | | | | |
| Tape Measure | 1 | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | | | | | | | | | | |
| Thermometer | 1 | | 1 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Tire Jack | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | |
| Trowel | | | | | | | | | | | | | | | | | 1 | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Tweezers | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | | 1 |
| Veggie Dicer | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | | | | | | | | | | | | |
| Washboard | | | | | | | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Watch | 1 | | | 1 | | | | | 1 | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Whisk | | 1 | | | | | | | 1 | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | 1 | | | 1 | | | | | | | | |
| Wire Cutters | | | | | | | | | 1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Table 4.3: The brainstorming outcomes (task 2) of different groups reveal the regularity and diversity human thinking process.
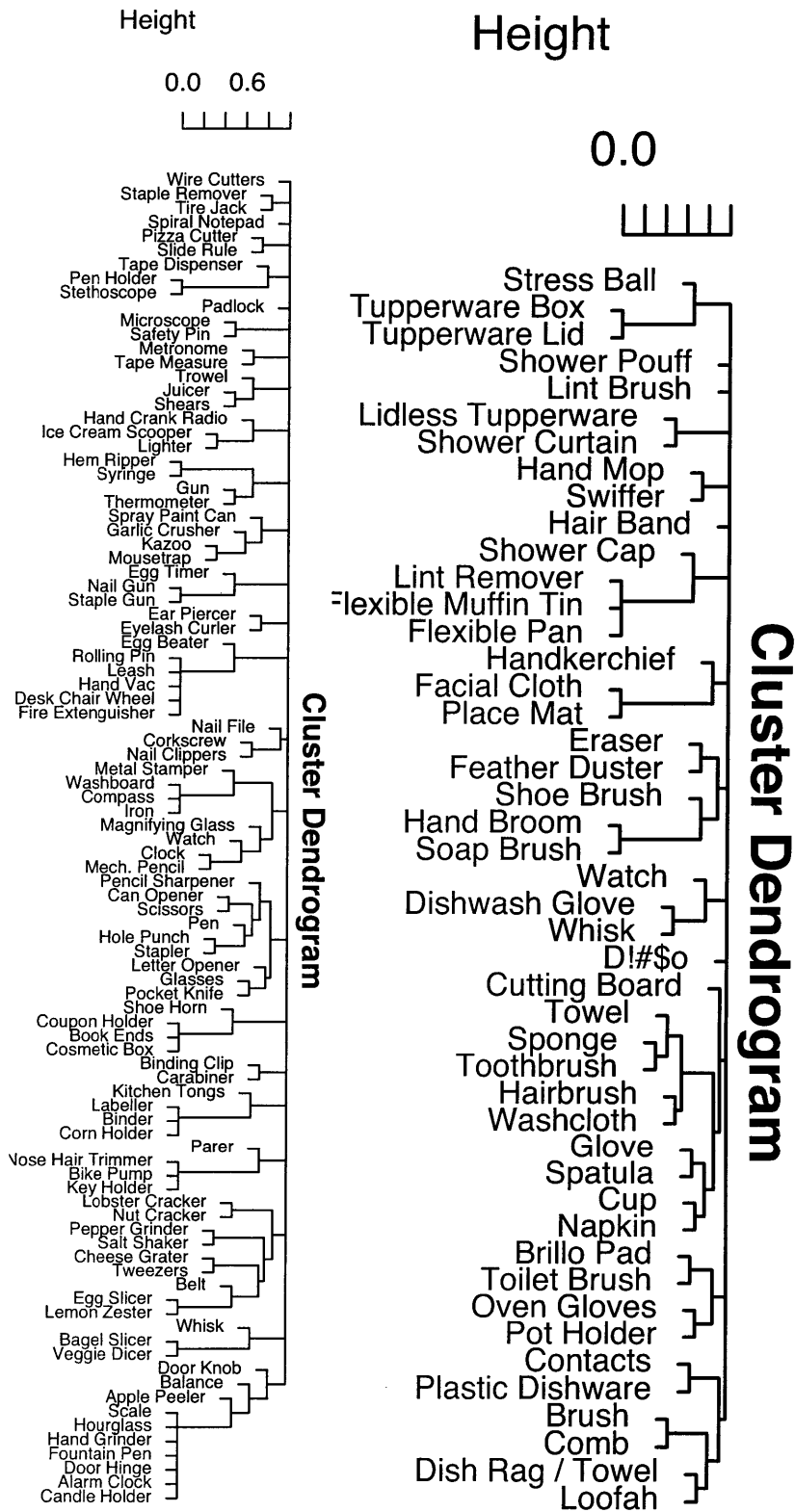
Table 4.4: ConceptNet indicates brainstorming has focus around familiar ideas (task 1).

| | brainstorming 1 ideas |
|---|---|
| bathroom | comb, cup, hairbrush, shower curtain, toilet brush, toothbrush, towel, washcloth |
| cabinet | brush, cup, eraser, glove, napkin, toothbrush |
| kitchen | cup, spatula, towel, washcloth, whisk |
| clean | brush, sponge, towel, washcloth |
| drawer | comb, eraser, spatula, whisk |
| home | cup, hairbrush, napkin, toothbrush |
| brush hair | brush, comb, hairbrush |
| closet | cup, towel, watch |
| groom | brush, comb, hairbrush |
| hair | brush, comb, hairbrush |
| hand | cup, glove, watch |
| house | cup, hairbrush, napkin |
| restaurant | cup, napkin, towel |
| shower | shower curtain, towel, washcloth |
| store | cup, eraser, toothbrush |
| straighten hair | brush, comb, hairbrush |
| suitcase | comb, glove, toothbrush |
| table | cup, napkin, place mat |
| taste good | eraser, shower curtain, toothbrush |
| this | comb, cup, eraser |

the subsequent questions to ask in the next session in order to finish the task. The following lists several strategies that a group can take in exploiting its brainstorming outcome.

A simple-minded approach is to assume that the brainstorming outcome contains the answer, and to conduct a "binary search" on all ideas generated in the brainstorming session. The shortcoming of this approach becomes apparent when the answer turns out to be outside of the brainstorming outcome: The information contained in the additional question/answer pairs is vague, and it needs a considerable amount of efforts to be processed — If the thing is neither a donut nor a bean, it could be not square, not edible, hard, larger than a golf ball, and so on.

A better strategy is to treat the brainstorming outcome as a skeleton of the space of all items compatible with the partially finished game, and to treat the compatible items outside of the brainstorming outcome as the flesh that grows on the skeleton. Thus the question of the group is not "is the thing either a donut or a bean?", but rather "is the thing edible?" randomly chosen among many choices. The improved strategy has the same performance as the simple-minded one when the answer is in the brainstorming outcome and it implies better performance when the answer is out of the brainstorming outcome.

A group can be more advanced-minded to estimate the probability measures that the group and the answerer respectively uses. Thus the group can discuss intensively on the consequences of asking "is it edible?" or "is it larger than a golf ball?" among many others before it finalizes the question.

We will discuss the decision-making performance of a group by two cases: when the answer is in the set of brainstorming ideas of the group, and when the answer is not in.

Based on our analysis, when the answer is in the set of brainstorming ideas, a group can identify the answer with around $\log_2 |S|$ questions (following the partially completed game), where $S$ is the set of brainstorming ideas. In the meeting mediator data set, 64% and 57% of the groups identified the correct answers with $1 + \log_2 |S|$ questions in task 1 and task 2 respectively, and 70% and 84% of the groups identified the correct answers in $2 + \log_2 |S|$ steps.

When the answer is not in the brainstorming ideas, the group will use around $\log_2 |S|$ steps to realize this fact, and then it will use additional $d(S, n)/|S|^{0.5}$ number of steps to reach the answer (cf., Fig. 4.2).

## 4.3   Individual Participation Levels & Group Performances

We have discussed how a 20-questions game is played by either an individual or a group, and how the performance at playing a game can be estimated from the result and performance of the preceding brainstorming session. We will continue to discuss how the group performance at playing such a game is determined by the ways that the group members act and interact.

In a 20-questions game, we have treated the answer space of a partially finished game — a list of yes/no questions and their answers — as a semantic network in which a node represents an idea compatible with the partially finished game, and the length of an edge connecting two nodes represents the difficulty to go *directly* along this edge to hit one another idea. Based on this treatment, a brainstorming process is a diffusion process in which the set of considered ideas spread along the edges on the semantic network, and a speaking-turn is stochastically an attempt to go along an edge. The statistics of a speaking turn can subsequently be related to the likelihood that a new idea or a new cluster of ideas is hit by the group: An isolated sub-second utterance ("yeah", "right", "OK") is unlikely a hit of an idea; A loud speaking turn followed by overlapping speaking-turns is more likely an indicator of a good progress; A cluster of loud speaking-turns followed by overlapping speaking-turns corresponds more likely to a cluster of ideas; And speaking turns are generally associated with the enumerating and testing of ideas.

Bales used the interaction process analysis (or IPA) to study the roles and communication pattern in a small task-group [Bales, 1950]. According to Bales, a task-group process is interwoven with task area roles (such as information giver, information seeker, orienteer, and follower) and socio-emotional area roles (such as protagonist, supporter, attacker, and neutral). In our work to detect these roles from signals that can be reliably estimated with automated tools, we found that the speaking/non-speaking time series of an individual is a SPLOT-type (stationary-process, large observation time) process [Trees, 2002], the time series from different individuals interact with each other by means of a latent combinatorial space, and there is a hierarchical structure in the group process [Dong et al., 2007]. Our findings from the perspective of digital signal processing coincide with those of Burke from a linguist's point of view [Burke, 1974]. We found it helpful to relate the statistics of roles and communication patterns to group performances.

The performance of a task-group can be estimated from the signals that are directly related with problem-solving, e.g., the brainstorming ideas generated by a group as we discussed in section 4.2.
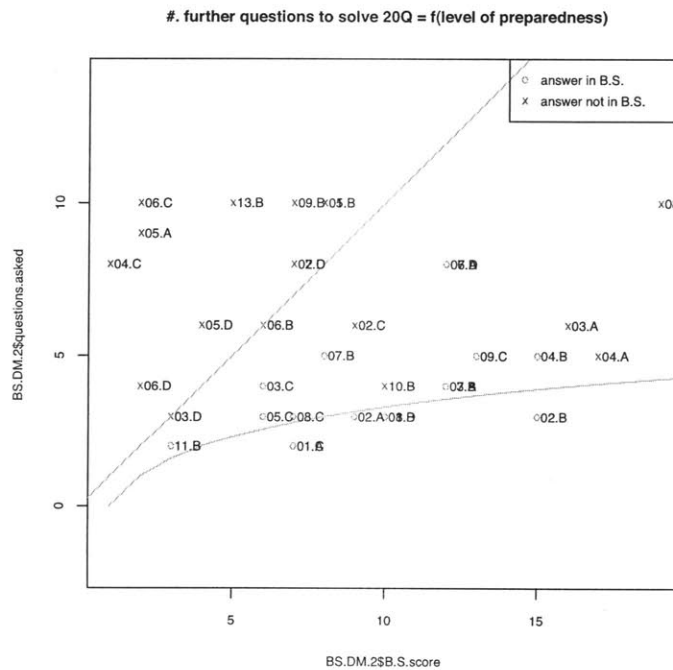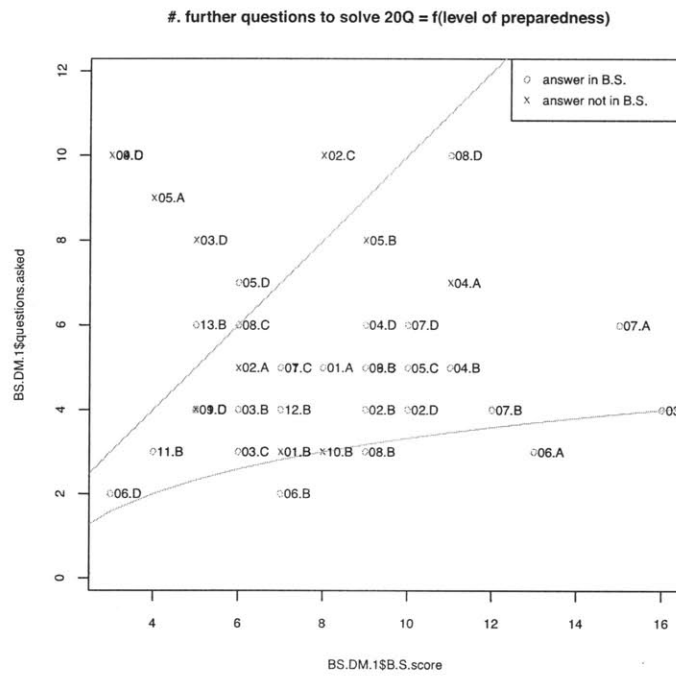
Figure 4.2: The number of questions asked by a group to solve a 20-questions game is a function of both the number of ideas generated in the preceeding brainstorming session and the structure of the ideas (task 1 and task 2). If the group hit the answer in brainstorming, the number of questions is approximately the logarithm of the number of ideas. If the group did not hit the answer, the number of questions is generally larger than the number of ideas. .

The performance can also be estimated from the signals that are indirectly related with problem-solving, such as the turn-taking behavior, speaking-turn lengths, audio amplitude and body movements of the group members. The former type of signal contains more information about the problem-solving, but it is much harder to collect especially with automated tools in real time. As a result, the meeting mediator is designed to estimate group performance and facilitate group collaboration based on the latter types of signals.

We should put a caution before we discuss the relationship between group process statistics and group performance, that we are dealing with statistics about human behavior and that human behavior is highly unpredictable. For example, a task group may effectively discuss a wrong topic, or a task group may consist of ingenious and socially-awkward people. In both of the two cases, the estimate of group performance based on group process statistics is wrong. Due to this caution, we build our group process model from the turn-taking signals all the way down to the signals directed related to the problem solving, and hope to make it clear in which sense we say that group process statistics and group performance are related.

The brainstorming and decision-making processes in the meeting mediator experiments are representative of many individual/group problem-solving processes. For example, a chess player (or a group playing one side) needs to brainstorm many possible sequences of moves, and make a decision on the next move based on the estimated consequences of each sequences. We hope that the problem-solving structure and the turn-taking structure of the 20-questions game could be related to those of other problems.

We used the speaking/non-speaking time series corresponding to the group members derived from the meeting mediators. These signals are then fed into an influence model to get the corresponding role assignment time series. The statistics that we can think of are subsequently computed and tested for their predictability of group performances.

Among the group process statistics that we tested, the *amount of speaker-overlap* is the best linear predictor of the group brainstorming performance in terms of either the total number of ideas or the total number of unique correct ideas generated by a group. ($R^2 \geq .33, p < .001$ for task 1 brainstorming and $R^2 \geq .35, p < .001$ for task 2 brainstorming.) This can be explained by our observation that a group uses most of its time in enumerating and testing ideas, and that our human subjects have comparable skills.

The sum of the fractions of time that the group members take the *giver role* or the *protagonist role* during a group brainstorming session is also a good linear indicator of the group brainstorming performance. This means that the brainstorming performance is explained by the long speaking-turns, which are supposed to be the places where the enumeration and validation of ideas really happen.

All the linear regressions in the above two paragraphs were carried out after we had taken out two data points (out of a total of around 40 data points for each regressions): One group had an average number of 2.5 simultaneous speakers during each of the 8 minute brainstorming sessions, and it only came out around 5 ideas and around 3 unique correct ideas in each sessions. Another group had an average number of 0.4 simultaneous speakers, and it came out around 32 ideas and around 16 unique correct ideas in each sessions. As a comparison, an normal group had an average of 1.2 simultaneous speakers, and it came out around 16 ideas and around 8 unique correct ideas in

each sessions. Due to the lack of audio/video recordings, we do not know whether the two group processes proceeded in the same way as the other group processes in the data set.

Leaders are reported to have impact on the task group performance. Due to our need to investigate the characteristics and roles of leaders, each group was instructed to choose a leader after a practice task that precedes and is similar to task 1 and task 2. The group was told that only its leader can ask questions in the following decision making sessions on the behalf of the whole group.

The leaders in the meeting mediator data set generate more ideas than the other group members in around 50% brainstorming sessions. They generated around 7.8 ideas on average for each brainstorming sessions, while a normal person generated around 4.5 ideas. and speak more and longer. The leaders spoke, took the information giver role and took the protagonist role more than the other group members in about 55% brainstorming sessions. Their speaking and turn-taking tendencies are consistent in both of their brainstorming sessions. Thus if we believe that the human subjects' performances are consistent in the practice task, task 1 and task 2, we can conclude that a task group pick its leader from those who perform well and participate well.

## 4.4 Improving Group Performance with the Meeting Mediators

We have discussed the structure of the idea space of a partially finished 20-questions game in the meeting mediator data set, the implications of this structure to the subsequent decision making sessions, and how the actions and interactions of the group members are related to the brainstorming performance of this group. We will proceed to discuss how the meeting mediators improved the group performances.

By encouraging interaction and engagement, the meeting mediators made noticeable improvements in regularizing the group behavior in the task-group processes without sacrificing performance, and the regularization pays off in performance when the task becomes harder. The more regularized and predictable group behavior has merits in management. In the experiments, the groups equipped with meeting mediators have their performances more linearly-predictable from the interactions in them ($R^2 \approx .40$ for groups with meeting mediators and $R^2 \approx .20$ for groups without, cf. Fig. 4.3). Indeed in performing task 2, even the groups not equipped with meeting mediators intrinsically had more interactions among their members, and their performances are more predictable from the amount of interactions in them. Recall that the overall R-squared statistic increases from .32 to .35 when the groups coped with the harder task (task 2). The meeting mediators also made the group members more considerate and the decision making faster, as was discussed by Kim et al. [Kim et al., 2008].

A new version of meeting mediator is under development that systematically exploits the group-process mechanism that we discussed and that could significantly improve the task-group performance.

**(a) Task 1 discussion intensity vs. outcome**
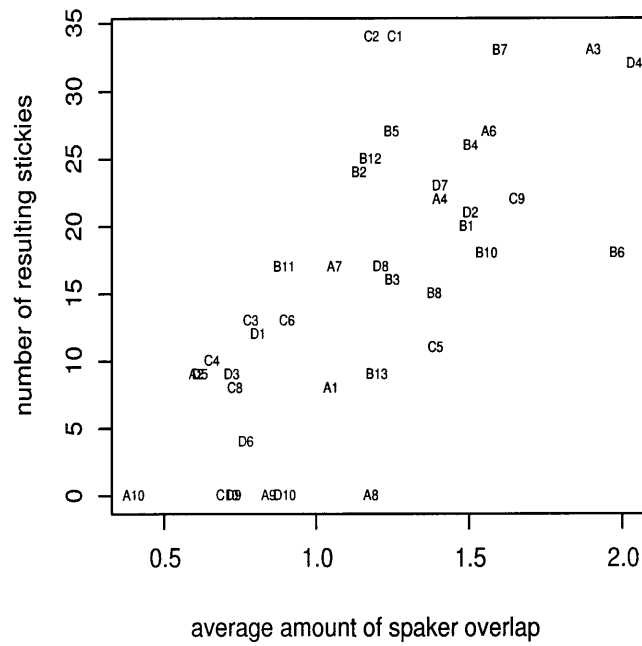


**(b) Task 2 discussion intensity vs. outcome**
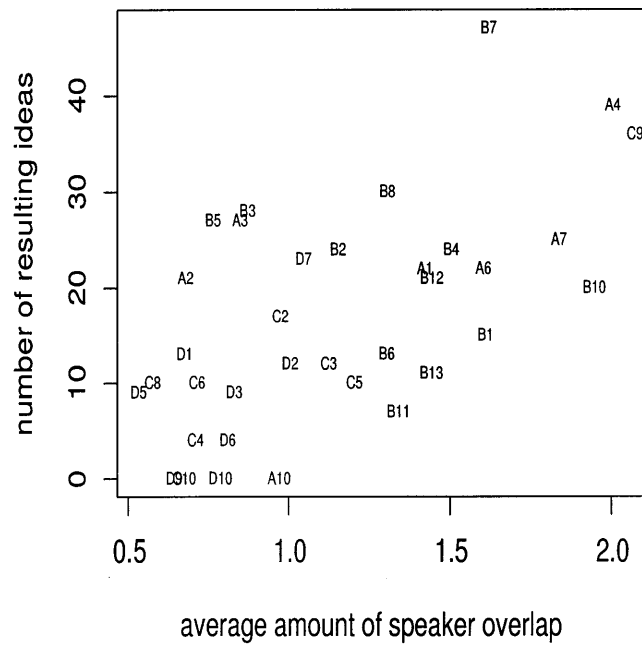


Figure 4.3: Brainstorming outcome vs. discussion intensity.

Table 4.5: ConceptNet indicates brainstorming has focus around familiar ideas (task 2).

| | brainstorming 2 ideas |
|---:|---|
| drawer | clock, corkscrew, egg beater, fountain pen, glasses, gun, letter opener, nail clippers, nut cracker, pen, pencil sharpener, stapler, tape dispenser, thermometer, whisk |
| desk | clock, fountain pen, gun, hole punch, iron, letter opener, pen, pencil sharpener, staple remover, stapler, tape dispenser |
| cabinet | compass, glasses, metronome, nail clippers, pencil sharpener, scale, staple remover, stapler, thermometer, trowel |
| this | alarm clock, balance, bike pump, hourglass, pen, pocket knife, safety pin, scale, stethoscope |
| office | clock, gun, letter opener, pen, pencil sharpener, staple remover, stapler, tape dispenser |
| tool | gun, iron, lighter, nail gun, stapler, stethoscope, tape measure, trowel |
| office supply store | binder, letter opener, pen, pencil sharpener, staple remover, stapler, tape dispenser |
| desk drawer | fountain pen, letter opener, pen, pencil sharpener, staple remover, stapler |
| pocket | fountain pen, glasses, lighter, pen, pocket knife, watch |
| school | binder, gun, pen, pencil sharpener, staple remover, stapler |
| home | alarm clock, clock, door hinge, glasses, stapler |
| library | binder, clock, pencil sharpener, stapler, tape dispenser |
| machine | belt, clock, metronome, scale, watch |
| office build | clock, door hinge, pen, stapler, tape dispenser |
| backpack | binder, can opener, compass, pen |
| doctor | scale, stethoscope, syringe, thermometer |
| doctor's office | scale, stethoscope, syringe, thermometer |
| house | clock, iron, pen, stapler |
| kitchen | can opener, iron, juicer, whisk |
| room | glasses, pen, scale, stapler |
| shop | pen, safety pin, stapler, tape measure |
| table | clock, pen, salt shaker, stapler |
| alarm | alarm clock, clock, watch |
| apartment | door hinge, pen, stapler |
| break | alarm clock, glasses, gun |
| build | pen, scale, stapler |
| car | clock, iron, lighter |
| city | gun, pen, stapler |
| classroom | pen, pencil sharpener, stapler |
| clock | alarm clock, hourglass, watch |
| device | stapler, stethoscope, thermometer |
| device measure time | clock, hourglass, watch |
| hold hand | gun, pen, staple gun |
| horse | compass, pen, watch |
| hospital | pen, syringe, thermometer |
| keep time | clock, metronome, watch |
| measure passage time | clock, hourglass, watch |
| measure time | clock, hourglass, watch |
| metal | gun, iron, safety pin |
| plastic | belt, pen, tape dispenser |
| post office | scale, stapler, tape dispenser |
| purse | nail file, pen, safety pin |
| store | clock, pen, stapler |
| time | clock, metronome, watch |
| useful | abacus, pocket knife, safety pin |
| work | clock, pen, stapler |

# Chapter 5

# Quantifying Group Problem Solving using Social Signal Analysis

Quantifying the relationship between group dynamics and group performance is a key issue of increasing group performance. In this chapter, 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. While we confine our discussion to a specific data set, we hope our findings and method could be applied to wider scenarios.

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 [Dong et al., 2009, Pentland, 2008]. We accomplish this by instrumenting group participants using *Sociometric Badges* [Olguin et al., 2009, Pentland, 2008], 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)* [Bales, 1950] 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* [mci]. 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.

I will summarize some key feature extraction steps, discuss some heuristics about quantifying group problem solving, and give our stochastic modeling that learns the structure of group problem solving.

## 5.1 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 (c.f. Fig. 5.1). The parameters of the hidden Markov model were set according to the manually marked task boundaries for three groups that we chose.

Figure 5.1: map button presses to task boundaries with Viterbi decoding.

We aligned the local times of different badges used in a same group process through the prin-
ciple 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
(c.f., Fig. 5.2).
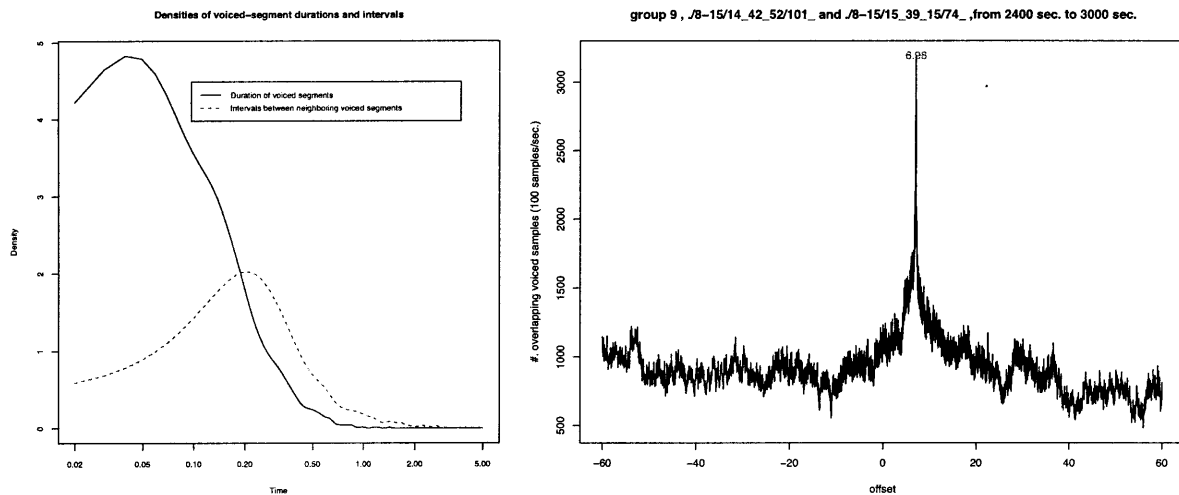


Figure 5.2: Left: The voice segments in our data set normally last 0.05 second and no longer than 0.5
second; They are normally 0.2 seconds apart when appearing in the same sentence. Right: Aligning voiced
frames could be a robust way to align data collected by different embedded devices deployed in 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 [Boersma, 1993],
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 inten-
sities 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.
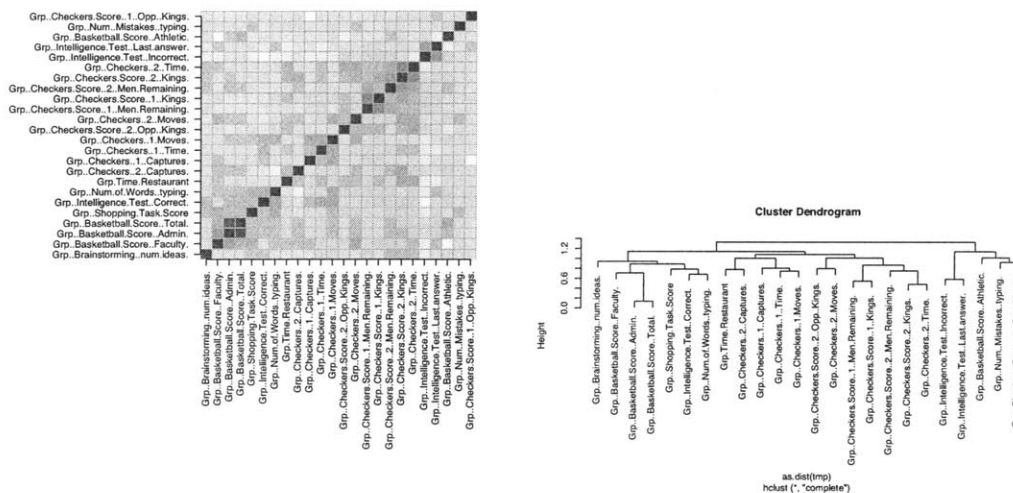
Figure 5.3: collective intelligence

## 5.2 Performance Heuristics

In this section, we show the correlations of the group performances in solving different problems in the MCI experiment, and the correlations between the group performances and group dynamics. The former class of correlations enable us to predict the performance of a group in solving a problem based on the performance of the same group in solving a similar problem. The latter class of correlations explain why one group performs better than another group.

The problem-solving skills involved in the MCI data set are organized in many clusters, as is indicated by Figure 5.3. The figure contains a *distance matrix* of the MCI problems and the corresponding *cluster dendrogram* constructed using the *complete linkage* method. In the distance matrix, the rows and columns are indexed by the MCI problems, and an entry at row $i$ and column $j$ represents the *distance* between problem $i$ and problem $j$. We define the *distance* between problem $i$ and problem $j$ in the following way so that they are closer if the group performance in solving one problem is more positively correlated with the group performance in solving the other problem. If two problems are closer in distance, they may involve more similar skill sets and problem solving mechanisms. In specific, let the numeric scores of the $n$ groups in solving problem $i$ be $X_1, \cdots, X_n$, and the numeric scores in solving problem $j$ be $Y_1, \cdots, Y_n$. If we treat the scores as samples of random variables $X$ and $Y$ respectively, The correlation coefficient between X and Y is subsequently $\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$. It can be verified that $d(i,j) = d(X,Y) = \sqrt{1 - \rho_{XY}}$ is a distance metric, and $d(\bullet, \bullet)$ is the distance we use in Figure 5.3.

Based on Figure 5.3, solving an MCI problem involves both general intelligence and problem-specific intelligence: A group that performs well in one checkers game problem normally performs well in another checkers game problem, and a group that performs well in one basketball problem normally performas well in another basketball problem; On the other hand, a group that performs well in the brainstorming problem and/or the cluster involving the group IQ problem (i.e., shopping problem, typing problem, group IQ problem) normally performs well in several other problems.

Figure 5.3 also shows that the number of questions attempted by a group in the group IQ test has visible positive correlation with the number of questions incorrectly answered by the group, however, the former doesn't seem to be related to the number of questions correctly answered by the group.

It is a clever and simple approach to predict the group performance of solving a problem from the group performances of solving other related problems. This approach shortcuts the complex and information-rich group-solving dynamics under the hood, and leaves us with a few good variables (the group performances in solving other related problems) to manipulate. On the other hand, the approach has several limitations due to its simplicity. Firstly, while we could collect data about solving simple problems in laboratory settings and estimate the correlation among the performances, we may have significant difficulty in finding a mathematical tool to estimate the correlation between the performance in solving a real-world problem and the performance in solving laboratory-setting problems. Secondly, the approach tells us nothing about how the group performances are related to the action and interaction of the group members and consequently how to facilitate group problem-solving. Thirdly, this approach doesn't enable us to tell the group performances in different parts of a group problem-solving, since the information related to the microscopic group dynamics is simply not available.

Non-linguistic cues could be used for performance prediction with appropriate parameterization suitable for the underlying process. Speaking-over, speech speed and equality of participation predict good momentum and thus good group performance in general. Sentence length and variability of volume predict group performance on a per task basis. For example, good group performance in analytical-reasoning corresponds to longer sentence length and less variability in volume than that in brainstorming.

Let us compare a process involving a group of three persons sitting together and solving an IQ test problem such as in the Raven's Progressive Matrices and another process involving a group in the same configuration and brainstorming the usages of a brick. A good performance in solving an IQ test problem normally involves clear reasoning and being able to promptly find the key patterns: Different individuals and groups normally have to spot the same key patterns and go through the same line of reasoning to get the right answer. As a result, a easy problem or a better group performance is normally related to the group dynamics involving one person speaking fluently and other two persons listening attentively. In comparison, a good performance in brainstorming is normally related to being able to act in fast tempo and think broadly. A long stretch of silence or a long stretch of speaking destroys the tempo and is normally not a good sign of brainstorming.

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.

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 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.

Fig. 5.4 gives the relationship between the number of clauses in an interaction process and the corresponding performance score in four tasks. This figure also gives subplots about the average numbers of clauses at different times of problem-solving for each individual processes. In this figure, there exist strong linear relationships between number-of-clauses and performance-score in the processes of the group brainstorming and group IQ-test. The number-of-clauses and performance-score relationship in solving the group IQ test may be slightly better fitted with a quadratic curve because the problems were increasingly harder and required more verbal reasoning in the processes. 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.

There are many task-specific heuristics: The brainstorming score of an interaction process in our data set is negatively correlated with the average clause length of this process. With support vector regression, the estimated brainstorming score from clause length could explain the brainstorming score with $R^2 = 0.45$. The number of back-and-forth interactions of group members in a group IQ test process could explain group IQ test score with $R^2 = 0.51$.

While it has many shortcomings, heuristic-based approach is widely-used to index group problem-solving performances, and it could be useful last-resort approach when other methods fail.
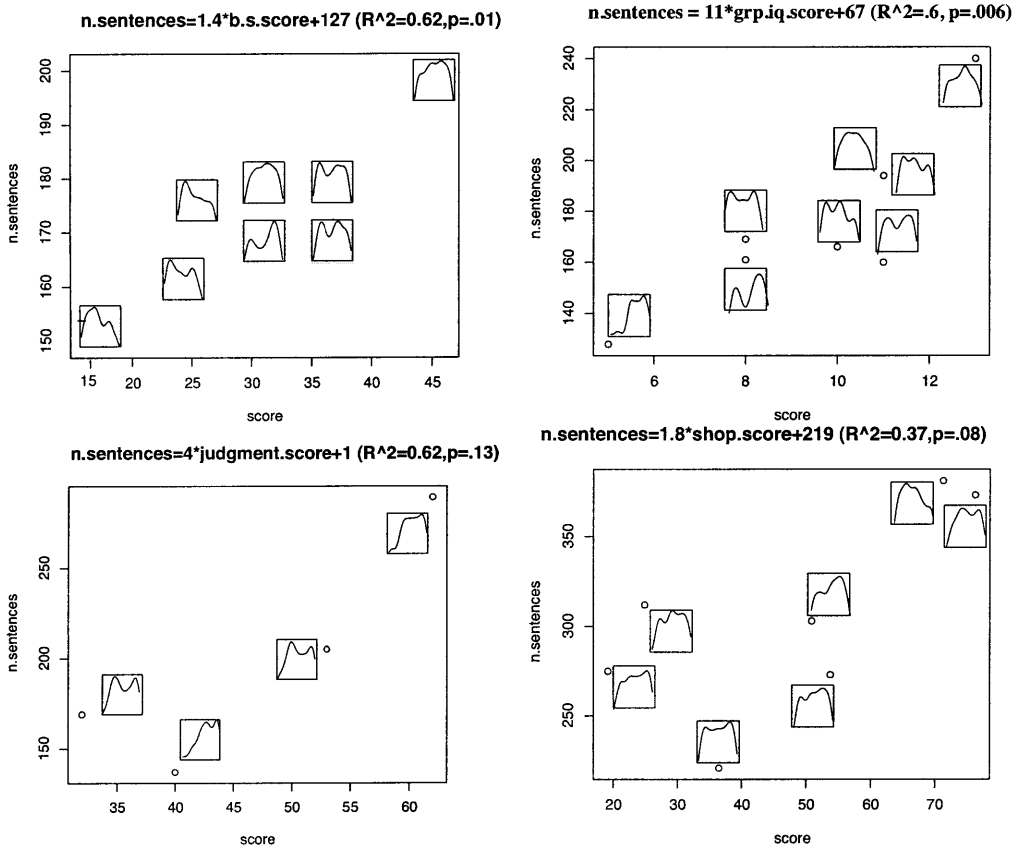
Figure 5.4: The number-of-sentences vs. performance-score relationship in four types of interaction processes: brainstorming, group IQ test, shopping and group judgment making (going clockwise from upper-left).

# 5.3 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 a mixture model of hidden Markov models (HMMs) to describe the probability measure of an interaction process, with each component HMM in charge of explaining the dynamics-performance relationship of solving each of the four specific types of problems.

There are similarities between the mixture of HMMs approach and how humans figure out the dynamics-performance relationships about group problem solving. Given a training set of interaction processes, together with the $n$ types of problems that the processes intended to solve and the performance scores, a human observer will intuitively relate the interaction processes to the corresponding problem types and performance scores. Hence he can 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 sentence 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 $n$ 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 model of hidden Markov processes, 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 $\theta_i$ 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 $\theta_0$ 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 $\theta_0$ 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 $(S_t, O_t)_{t=1\cdots T}$.

Of the 43 interaction processes in solving the four tasks that we are interested in, we have randomly chosen 80% processes for training and the rest 20% for testing. We could estimate performance with $R^2 \approx 60\%$ accuracy by finding the maximum likely performance covariate.

The performance coefficients of the fitted models for the four tasks (c.f. table 5.1) tell us not only

| | b.s. | grp.iq | shop | jdgmnt |
|---|---|---|---|---|
| $\mathbb{P}(s_1 \to s_1)$ | .9-f*2e-5 | .9-f*3e-5 | .9 | .9 |
| $\mathbb{P}(s_2 \to s_2)$ | .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 5.1: Hidden Markov modeling of interaction processes in solving four tasks.

how different tasks require different group process dynamics but also how different performances in the same task correspond to slightly different dynamics. In general good performance generally requires more 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 sentences ($\mu(\Delta t|s_1)$), 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 sentence 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 involves making good reasoning. The longer sentence 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.

This active-discussion and good-performance relationship could be further confirmed by estimating the heuristics at different performance levels through simulating the component HMMs (c.f. table 5.2). For 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 sentences and involve 4, 5, 7 speaker changes per minute.

In summary, it is generally not possible without prior knowledge to relate hours of signals of group problem-solving to limited amount of performance labels. Modeling the conversational interaction is a good step forward to bring in prior knowledge. We can proceed to bring in more prior knowledge about human problem-solving. But a better way is to try to align the dynamics with the performance based on data mining.

| percentile | number of sentences | sentences per person*minute | speaker. turns per minute | vowels per sentence | 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) A lookup table to estimate performance from observed dynamics in the brainstorming task in the MCI data set.

| task | 1 word | 2 words | 3 words | > 4 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) The sentence-length statistic is adapted to better task performance in the MCI data set.

Table 5.2: Hidden Markov modeling of interaction processes in solving four tasks.

# Chapter 6

# Conclusions and Further Discussions

In this thesis, I have demonstrated my approach to the modelling of the randomness and diversity of human problem solving. This approach uses influence modeling to track the dynamics and measure the effectivenss of problem solving by examining the signals that can be reliably collected with embedded devices. This method allows us to formulate and test models of human problem solving processes, and to explore possibilities for improving performance with a model-based, data-driven approach. The "common sense" knowledge that I used in this thesis either comes from experimental observations, or from a common sense knowledge database such as ConceptNet. It is my sincere hope that research into human intelligence can ultimately make people more intelligent, that we can collect enough data to sufficiently represent the diversity and randomness of human intelligence. I futher hope that we can one day study human intelligence with stochastic differential equations and datawith the same effectivenss with which physicists currently study the natural world.

From my perspective, a good problem solver is a resourceful person. The resourceful person has both an efficiently connected knowledge network and a flexible attitude when attemptingdifferent ways of reaching a goal. Problem solving may be conceptualized as finding a path from a given initial node to a goal node within a network. Hence, to facilitate problem solving, a person or a computer must have enough domain knowledge and an effective model of the problem solving process. My preceding perspective draws upon on my own experience, previous research, as well as folk wisdom about constructing tools to facilitate human problem solving. The ACT-R theory [Anderson, 1990, Anderson and Bellezza, 1993] provided me with experimental facts and an explanation how people both organize their experience and retrieve it during the stepwise solving of their problems. My participation in the programs of MIT Teaching and Learning Laboratory enables me to see how teachers in different academic fields engage and help students solve problems by putting knowledge in contexts relevant to the students. Many general works about problem-solving [Newell and Simon, 1972, Sternberg and Frensch, 1991, Frensch and Funke, 1995] in domains such as mathematics [Hadamard, 1954,1945, Pólya and Conway, 2004,1985], physical sciences, engineering [Mahajan, 2008], reading and writing[Kaye, 1989] have been written. I have read more than 100 works of this type. The study of such excellent works has provided me with broad background of solid information about how problem solvers actively explore their knowledge to solve their problems, as well as how previous observers, researchers and authors have conceptualized the problem-solving process. The common sense knowledge databases, such as

79

ConceptNet, have enabled me to fit random walk models into discourses, and to simulate how the thinking process wander among conflicting opinions and different moods. My research related to Honest Signals [Pentland, 2008] has helped me understand how verbal communication in a small-group problem-solving process is related to the (sub)conscious non-linguistic social signals displayed by the individuals, as well as non-linguistic interpersonal interaction patterns. It has also given me opportunities to construct stochastic processes to make inferences about content from non-linguistic social signals [Dong, 2006, Dong and Pentland, 2007].

The ACT-R cognitive architecture [Anderson, 1990, Anderson and Bellezza, 1993], is rooted in the General Problem Solver [Newell and Simon, 1972]. This work painstakingly attempts to explain the macroscopic performance of problem solvingin terms of the amount of time and the earned score of finishing a task, from the microscopic behavior of how individuals access their memory and make stepwise decisions. ACT_R does this in terms of the memory access times and the estimated gains corresponding to alternative steps used by a person in order to achieve their goals of problem solving. Over a period of several decades, researchershave used the cognitive architecture of ACT-R to explain observations made in the laboratory of many well-controlled problem-solving experiments (e.g., solving a Hanoi tower problem, planning a trip under uncertain conditions, making predictions about new observations) and learning (e.g., learning LISP programming language). I have found the several aspects of ACT-R to be useful. For instance, the hard facts provided about the experiments have been valuable., The use of microscopic behavioral dynamics to explain macroscopic performance, which parallels the approach commonly used by physicists, has been very useful. The rational analysis approach, which biases stepwise decision-making towards the direction that attempts to maximize performance, has likewise proved quite useful On the other hand, many formulae in ACT-R lack simplicity and physical explanation. There seems to be no information provided about how parameters of the ACT-R formula were fitted to the experimental data. Furthermore, I have significant difficulties in applying ACT-R to more complex problem-solving instances, such as those involving natural languages and collaboration.

The teachers in various academic fields seem to feel they can best assess a students' mastery of knowledge and facilitate each students' problem-solving based on introspectively pondering how students think, based solely on their own "intuitions" and implicit knowledge. This approach is different from that of some managers and politicians, who resort to oversimplified slogans such as "brainstorming is always an effective way of problem solving," and "let us ask the primary school students to construct mathematical knowledge all by themselves since constructivism is right." I believe the teachers' approach is a better approach to helping people. Further, it is my intention to contribute to building a comprehensive database of diverse human knowledge and provide a stochastic description of how people access their own knowledge to solve problems. Doing so may allow us to transform the "black art" of teaching, based solely upon the implicit knowledge, into a quantitative science, based on knowledge, databases, measuring and simulation. My understanding of what teachers do comes from my own experience, and from my participation in the programs of the MIT Teaching and Learning Laboratory (TLL). By my understanding, TLL advocates[Bransford et al., 1999, Prosser and Trigwell, 1999, Biggs, 2006] that learning is knowledge constructed within the minds of the students. \An expert differs from a novice in that experts have a more efficient network of concepts and their relations acquired from past experience than do the novices. According to this view, teaching and learning should be evaluated in the in terms of increased ability of the student to engage their knowledge for use in problem solving

This approach tolearning and problem-solving seems to agree with the perspective of Pólya [Pólya and Conway, 2004,1985], who primarily concerned withsolving math problems. In*How to solve it: a new aspect of mathematical methods*, Pólya pointed out that the role of a teacher is to guide the students - without giving them too many hints - to extract the meaning of the problem, the relevants details about the related concepts and analogous problems and solutions. Polya did this so that the students could build a path from the information given to the goal by themselves. This perspective on learning and problem-solving also seems to agree with that of Minsky [Minsky, 1988,1986, 2006], Lenat [Lenat and Guha, 1989,1990] and others, who suggest we should give a computer a large amount of knowledge so that the computer can reason like a human, and can adaptively assist people in their everyday lives. It is my hope that the stochastic description of the problem/solution developed in my research can eventually assist people in reasoning about problem solving with existing knowledge databases, and can encourage people to contribute to the creation of a comprehensive knowledge database that captures the diversity of our society.

Introspection upon my own experience seems to confirm the idea that thinking is a random walk in the direction of intention or problem solution, drawing upon on past experiences which form a knowledge network. Imagine for instance, a brief experiment in which several students in a lecture are given one minute to sum the first $n$ odd natural numbers, with the teacher being responsible for facilitating the students' problem-solving. Most students might immediately speculate the solution to be $n^2$ by following pattern that the sums corresponding to $n = 1, 2, 3, 4$ are respectively 1, 4, 9, 16. Most students might proceed to confirm their speculation, but some might fail to do so. Many students would proceed to consider the other approaches to solving the problem.. Some would stack up rows of 1, 3, 5, 7 square blocks and count the volume of the resulting triangles, and this approach algebraically corresponds to $\sum_1^n (2k - 1) = \frac{1}{2} \sum_1^n ((2k - 1) + (2n - (2k - 1))) = n^2$. Some would repetitively put 3, 5, 7 square blocks on two neighbouring edges of the $1 \times 1$, $2 \times 2$, $3 \times 3$ block squares to get larger block squares, and this approach algebraically corresponds to $\sum_1^n (2k - 1) = 1 + \sum_2^n (k^2 - (k - 1)^2) = n^2$. Other students mightreduce summing the first $n$ odd numbers to summing the first $2n$ natural numbers.

I have the following comments to make about common sense knowledge concerning the solution by special cases approach in the preceding example. Solution by special cases naturally came to the students because it was developed in an early stage of mental development and became common sense, but it may be forgotten when people cope with complex problems. It worked in the example given above because it is "common sense" that 1, 4, 9 are square numbers. The conception of proving the truthfulness and disproving the falsity of a statement has been taught and naturally held by most science and engineering students, but this conception may not be obvious to everyone.

Concerning the deeper and more abstract approaches related to the preceding example, I hold the perspective that abstract-thinking involves the abilities of freely navigating between a generalization and its many specific examples, finding commonality, making connections among seemingly unrelated things, and error-free reasoning. This perspective agrees with the constructivism perspective advocated by TLL, which claims that problem- solving can be facilitated by a faster access to the appropriate context than is enabled from the past experience. To elaborate my perspective in the example, I can see no reason that the students chose the deeper approaches among many alternative ways of operating upon the odd numbers (e.g., decomposing $2k - 1$ into $k - 1$ and $k$) if they had not had previous experience with triangles and squares. I can see no superiority of a

human problem solver over a computer — which has no human experience, but is programmed to enumerate all possible ways of combining odd numbers — either, if previous human experience does not play a role.

Bransford et al. [Bransford et al., 1999] described many examples of problem/solution and context/experience. Of these, I found the psychological study by de Groot on expert chess playing to be the most interesting. In this example, the chess masters and the novices were found to have equally good recall when faced with random arrangements of chess pieces. However, when presented with strategically-meaningful arrangements of pieces, the recall of chess masters was vastly superior to that of novices. The masters had more patterns in stored within their memories from past experience, and this enabled them not only remember the chess positions taken from actual games, but also to anticipate the moves of opponents, and hence perform better than the novices. I believe that problem solution without context could be extremely difficult, and we should not judge the ability of abstract thinking by asking people to solve problems without context.

In many important applications of natural languages, such as sharing opinions and coordinating problem-solving efforts in both written and spoken forms, we perform random walks in the networks of our individual past experience in the directions that maximize our values. In such networks, every node represents a piece of experience and has a value. These nodes are all connected to one another according to how "close" they are. If we somehow collect a comprehensive network of human experience, the network can be applied to model the dynamics and interpersonal "influences" in brain storming, group problem solving, writing and reading from observations. It can also model all discourses as samples of the corresponding opinions based upon Markov chain Monte Carlo.

The network random walk perspective can be considered when studying writing. As reviewed by Bryson and Bereiter [Sternberg and Frensch, 1991], the comparisons based on thinking-aloud protocols seemed to suggest the expert writers tended to use the larger pictures of extracting and effectively organizing arguments in conveying the experts' actual opinions to the audience. They also note it is possible to make novice writers perform like experts by instructing the novices to cope with the more important problems of finding out the truths from past experience and engaging the audience. The differences between an expert and a novice in writing is consistent with the differences in other areas of problem solving. The more efficient problem solution behavior of an expert can be explained by the fact that their additional experience contributes to their more efficiently-connected knowledge network, which in turn enables them to navigate their network in more efficient ways [Bransford et al., 1999]. This concept that writing is a problem/solution and requires the engaging past of experience seems to be endorsed byseveral experts. For example, Kaye [Kaye, 1989] gave a timetable about how to engage experience and finish a 3-page article in three hours.

I hope one day that it will be possible to transform the study of the human intelligence into a field akin to a physical science, in which we make good predictions about people based on various stochastic equations and sufficiently digitized human experiences. I propose to treat human thinking as a random-walk in a network of human knowledge. This treatment reminds us that we can sometimes oversimplify problems and make biased judgments based on our own limited experiences. It also provides us with a good mathematical foundation to quantitatively discuss the microscopic details about the diversity and randomness of human problem solving processes. Our

behavior is limited by our experiences, and it the behavior of a computer. When I imagine myself as a present-day computer participating in the human-computer interaction beyond its programmed capability, I realize that the computer is either blank in its "thinking" or is inclined to ask for further instructions. The computer has no intentions of its own, no tidbits of human experience to search for a solution among, and no clues about how to speculate upon the mental status of its human partner. Thus, at best, it acts like a philosophical zombie.

When we abstract the content of the group problem-solving process and focus on the form (e.g. the interaction pattern among the participants), the psychological theories, such as Bales' interaction process analysis (IPA), make the stochastic analysis easier by imposing a structure on the process. In Bales' IPA, a group problem-solving process is composed of a series of events involving periods of providing/seeking facts, providing/seeking opinions on the facts, suggesting/seeking approaches for problem solving, voting for different approaches, and solving socio-emotional conflicts. Hence, IPA summarizes group problem-solving by counting the number of events of different types. From the perspective of a computer scientists, it enables the algorithmical estimation of group discussion performance by translating gigabytes of information sampled by embedded devices in the group process into a process of Bales' events, as demonstrated in this thesis. As I understand it, Bales' general form of group problem-solving process should be related to the content and verified based on common sense knowledge databases, such as ConceptNet. Since any abstract form only makes sense in the context where it can be substantialized with content, this should be done for future studies. . There are two reasons for studying the form of a process: (a) the goal of training people in problem-solving is to enable them to be good problem-solvers in future scenerios that are different from the training scenarios; and (b) our current technologies are still unable to reliably estimate the content of group problem-solving.

I would like to thank several professors for advising me to point out the limitations of making statistical inferences with forms and without contents. This thesis only tells my observations about the statistical correlations between the dynamics and the performances, investigates these correlations by relating the form to the content, and suggests people adjust their dynamics and to improve their performances based on my limited observations. It should neither be taken as a way to "fake" good performance nor as an absolute measure of performance.

The theory of honest signals [Pentland, 2008] represents a novel approach to abstracting the content of individual and organizational behavior, and for investigating the relationship between performance and dynamics generally. According to this theory, we should put more emphasis on the research of pre-linguistic social signals (such as the centrality role of a person in a social network, the way they communicate with others, their psychological profile, and, at least from my perspective, their habits ). This is because these social signals form the basis of human language. This thesis showed my attempt to interpret the honest signals in terms of performance data and the linguistic cues I could collect in the experiments. The field of artificial intelligence (AI), which aims to simulate human intelligence with computer programs, started when the first digital computers came into being as an attempt to create a computer with superhuman intelligence Failing that, AI eventuallyproceeded to the deceptively difficult problem of merely creating acomputer that displays what seems to humans like "common sense." [Minsky, 1988,1986]. In my opinion, the field should proceed even further, and equip a computer with the honest signals. What we can achieve and understand are limited by our experiences, and this is equally true of computers.

# Bibliography

A description of the twenty questions game. URL http://en.wikipedia.org/wiki/20_ questions.

MIT Committee on the Use of Humans as Experimental Subjects. URL http://web.mit. edu/committees/couhes/index.shtml.

Measuring Collective Intelligence Studies. http://cci.mit.edu/research/ measuring.html.

John R. Anderson. *The adaptive character of thought*. L. Erlbaum Associates, Hillsdale, N.J., 1990.

John R. Anderson and Francis S. Bellezza. *Rules of the mind*. L. Erlbaum Associates, Hillsdale, N.J., 1993.

C. Asavathiratham, S. Roy, B. C. Lesieutre, and G. C.Verghese. The influence model. *IEEE Control Systems Magazine, Special Issue on Complex Systems*, (12), 2001.

Chalee Asavathiratham. *The Influence Model: A Tractable Representation for the Dynamics of Networked Markov Chains*. PhD thesis, MIT, 1996.

J. Maxwell Atkinson, editor. *Structures of Social Action (Studies in Conversation Analysis)*. Cambridge: Cambridge University Press, 1985.

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

Robert Freed Bales. *Personality and Interpersonal Behavior*. Holt, Rinehart and Winston, 1969.

Robert Freed Bales and Stephen P. Cohen. *Symlog, A System for the Multiple Level Observation of Groups*. Free Press, 1979.

Satanjeev Banerjee and Alexander I. Rudnicky. Using simple speech-based features to detect the state of a meeting and the roles of the meeting participants. In *Proceedings of the 8th International Conference on Spoken Languate Processing*, volume 2189-2192, 2004.

Regina Barzilay, Michael Collins, Julia Hirschberg, and Steve Whittaker. The rules behind roles: Identifying speaker role in radio broadcasts. In *Proceedings of AAAI/IAAI*, pages 679–684, 2000.

Sumit Basu, Tanzeem Choudhury, Brian Clarkson, and Alex Pentland. Learning human interactions with the influence model. Technical report, MIT Media Laboratory Vision & Modeling Technical Report #539, 2001. URL http://vismod.media.mit.edu/tech-reports/TR-539.pdf.

L. E. Baum, T. Petrie, G. Soules, and M. Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *Annual of Mathematical Statistics*, 41 (1):164–171, 1970.

John Biggs. Teaching teaching & learning learning. http://www.daimi.au.dk/~brabrand/short-film/, 2006.

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.

Xavier Boyen and Daphne Koller. Exploiting the architecture of dynamic systems. In *AAAI/IAAI*, pages 313–320, 1999.

Mathew Brand. Coupled hidden markov models for modeling interacting processes. Technical report, MIT Media Laboratory Vision & Modeling Technical Report #405, 1996. URL http://vismod.media.mit.edu/tech-reports/TR-405.ps.Z.

Mathew Brand, Nuria Oliver, and Alex Pentland. Coupled hidden markov models for complex action recognition. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1997.

John. Bransford, Ann L. Brown, and Rodney R. Cocking, editors. *How people learn : brain, mind, experience, and school*. National Academy Press, Washington, D.C., 1999.

Peter J. Burke. Participation and leadership in small groups. *American Sociological Review*, 39: 832–843, 1974.

Giliano Carli and Roberto Gretter. A start-end point detection algorithm for a real-time acoustic front-end based on dsp32c vme board. In *Proceedings of ICSPAT*, pages 1011–1017, 1997.

P. Chippendale. Towards automatic body language annotation. In *Proceedings of the 7th International conference on Automatic Face and Gesture Recognition*, pages 487–492, 2006.

Tanzeem Choudhury. *Sensing and Modeling Human Networks*. PhD thesis, MIT, 2003.

Tanzeem Choudhury and Alex Pentland. The sociometer: A wearable device for understanding human networks. *Computer Supported Cooperative Work - Workshop on Ad hoc Communications and Collaboration in Ubiquitous Computing Environments*, 2002.

M. Diehl and W. Stroebe. Productivity loss in idea-generating groups: tracking down the blocking effect. *Journal of Personality and Social Psychology*, 61:392–403, 1991.

Wen Dong. Infuence modeling of complex stochastic processes. Master's thesis, MIT, 2006.

Wen Dong and Alex Pentland. Modeling influence between experts. *Artificial Intelligence for Human Computing*, 4451(170-189), 2007.

Wen Dong, Bruno Lepri, Alessandro Cappelletti, Alex Pentland, Fabio Pianesi, and Massimo Zancanaro. Using the influence model to recognize functional roles in meetings. In *ICMI*, pages 271–278, 2007.

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.

Peter A. Frensch and Joachim Funke, editors. *Complex Problem Solving: the European Perspective*. Lawrence Erlbaum, 1995.

Lawrance R. Frey, Dennis S. Gouran, and Marshall Scott Poole. *The Handbook of Group Communication Theory and Research*. SAGE, 1999.

Neha P. Garg, Sarah Favre, Hugues Salamin, Dilek Z. Hakkani-Tür, and Alessandro Vinciarelli. Role recognition for meeting participants: an approach based on lexical information and social network analysis. In *ACM Multimedia*, pages 693–696, 2008.

Malcolm Gladwell. *The tipping point : how little things can make a big difference*. Little, Brown, Boston, 2002.

Jacques Hadamard. *An essay on the psychology of invention in the mathematical field.* Dover Publications, [New York], 1954,1945.

Jay Hall and W. H. Watson. The effects of a normative intervention on group decision-making performance. *Human Relations*, 23(4):299–317, 1970.

Hayley Hung, Yan Huang, Gerald Friedland, and Daniel Gatica-Perez. Estimating the dominant person in multi-party conversations using speaker diarization strategies. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2197–2220, 2008a.

Hayley Hung, Dinesh Babu Jayagopi, Sileye O. Ba, Jean-Marc Odobez, and Daniel Gatica-Perez. Investigating automatic dominance estimation in groups from visual attention and speaking activity. In *ICMI*, pages 233–236, 2008b.

J. M. Jackson and S. G. Harkins. Equity in effort: An explanation of the social loafing effect. *Journal of Personality and Social Psychology*, 49:1199–1206, 1985.

Dinesh Babu Jayagopi, Sileye O. Ba, Jean-Marc Odobez, and Daniel Gatica-Perez. Predicting two facets of social verticality in meetings from five-minute time slices and nonverbal cues. In *ICMI*, pages 45–52, 2008.

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

Sanford. Kaye. *Writing under pressure : the quick writing process*. Oxford University Press, New York, 1989.

Taemie Kim, Agnes Chang, Lindsey Holland, and Alex Pentland. Meeting mediator: Enhancing group collaboration and leadership with sociometric feedback. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, 2008.

Douglas B. Lenat and R. V. Guha. *Building large knowledge-based systems : representation and inference in the Cyc project*. Addison-Wesley Pub. Co., Reading, Mass., 1989,1990.

Sanjoy Mahajan. The art of approximation in science and engineering. `http://ocw. mit.edu/OcwWeb/Electrical-Engineering-and-Computer-Science/ 6-055JSpring-2008/Readings/index.htm`, 2008.

Iain McCowan, Daniel Gatica-Perez, Samy Bengio, Guillaume Lathoud, Mark Barnard, and Dong Zhang. Automatic analysis of multimodal group actions in meetings. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27(3):305–317, 2005.

J. McGrath and D. Kravitz. Group research. *Annnual Review of Psycology*, 33:195–230, 1982.

J. E. McGrath. *Groups: Interaction and Performance*. Prentice-Hall, 1984.

Marvin Lee Minsky. *The society of mind*. Simon & Schuster, New York, 1st touchstone ed. edition, 1988,1986.

Marvin Lee Minsky. *The emotion machine : commensense thinking, artificial intelligence, and the future of the human mind*. Simon & Schuster, New York, 2006.

Allen Newell and Herbert Simon. *Human Problem Solving*. Prentice-Hall, 1972.

Bernard A. Nijstad, Wolfgang Stroebe, and Hein F. M. Lodewijkx. Production blocking and idea generation: Does blocking interfere with cognitive processes? *Journal of Experimental Social Psychology*, 39:531–548, 2003.

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.

Nuria M. Oliver, Barbara Rosario, and Alex Pentland. A bayesian computer vision system for modeling human interactions. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 22(8), pages 831–843, 2000.

A. F. Osborn. *Applied Imagination: Principles and Procedures of Creative Problem Solving*. New York, NY: Charles Scribner's Sons, 1963.

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

Fabio Pianesi, Nadia Mana, Alessandro Cappelletti, Bruno Lepri, and Massimo Zancanaro. Multimodal recognition of personality traits in social interactions. In *Proceedings of the 10th International Conference on Multimodal Interfaces: Special Session on Social Signal Processing*, pages 53–60, 2008a.

Fabio Pianesi, Massimo Zancanaro, Bruno Lepri, and Alessandro Cappelletti. A multimodal annotated corpus of concensus decision making meetings. *Language Resources and Evaluation*, 41(3-4):409–429, 2008b.

George Pólya and John Horton. Conway. *How to solve it : a new aspect of mathematical method*. Princeton University Press, Princeton [N.J.], expanded princeton science library ed. edition, 2004,1985.

Michael Prosser and Keith Trigwell. *Understanding Learning and Teaching (Society for Research into Higher Education)*. Open University Press, 1999.

Rutger Rienks and Dirk Heylen. Dominance detection in meetings using easily obtainable features. In *MLMI*, pages 76–86, 2005.

Rutger Rienks, Dong Zhang, Daniel Gatica-Perez, and Wilfried Post. Detection and application of influence rankings in small group meetings. In *ICMI*, pages 257–264, 2006.

Harvey Sacks, Emanuel A. Schegloff, and Gail Jefferson. A simple systematics for the organization of turn-taking for conversation. *Languate*, 50(4), 1974.

Lawrence K. Saul and Michael I. Jordan. Mixed memory markov models: Decomposing complex stochastic processes as mixtures of simpler ones. *Machine Learning*, (37):75–86, 1999.

Emanuel A. Schegloff. *Sequence Organization in Interaction: A Primer in Conversational Analysis*, volume 1. Cambridge: Cambridge University Press, 2007.

Robert J. Sternberg and Peter A. Frensch, editors. *Complex Problem Solving: Principles and Mechanisms*. Lawrence Erlbaum, 1991.

Harry L. Van Trees. *Detection, Estimation, and Modulation Theory, Part III*. John Wiley & Sons, 2002.

A. Vinciarelli. Role recognition in broadcast news using bernoulli distributions. In *Proceedings of IEEE International Conference on Multimedia & Expo*, pages 1551–1554, 2007a.

Alessandro Vinciarelli. Speakers role recognition in multiparty audio recordings using social network analysis and duration distribution modeling. *IEEE Transactions on Multimedia*, 9(6): 1215–1226, 2007b.

Chung Y. Weng, Wei T. Chu, and Ja L. Wu. Movie analysis based on roles' social network. In *Proceedings of IEEE International Conference on Multimedia & Expo*, pages 1403–1406, 2007.

Darren James. Wilkinson. *Stochastic modelling for systems biology*. Taylor & Francis, Boca Raton, FL, 2006.

David Sload Wilson, John J. Timmel, and Ralph R. Miller. Cognitive cooperation: When the going gets tough, think as a group. *Human Nature*, 15(3):1–15, 2004.

Massimo Zancanaro, Bruno Lepri, and Fabio Pianesi. Automatic detection of group functional roles in face to face interactions. In *ICMI*, pages 28–34, 2006.