Towards a Computational Model of Small Group Facilitation

Yoichi Matsuyama∗
Perceptual Computing Group,
Waseda University
27 Waseda, Shunjuku-ku, Tokyo, 162-0044, Japan
matsuyama@pcl.cs.waseda.ac.jp

Tetsunori Kobayashi
Perceptual Computing Group,
Waseda University
27 Waseda, Shunjuku-ku, Tokyo, 162-0044, Japan
koba@waseda.jp

Abstract

A physically situated robot in conversational situations, which is enabled to understand and generate human conversational protocols, has a big potential to facilitate conversation in a small group. In this paper, we present a computational model of facilitation process in a small group, including (1) procedural behaviour decision process controlling engagement density to regulate a socially imbalanced situation and (2) language generation process associated with user models, which attempts to trigger participants' interests. We implemented the model on a conversational robot and assessed the effectiveness of the procedural behavior generation.

Introduction

While traditional research on embodied conversational agents (ECAs) have strong assumption to study dyadic interactions (Cassell and Bickmore 2003), many researchers presented differences between two-participant interactions and multiparty interactions, and dyadic models cannot be easily applied to multiparty cases. Numerous research on multiparty interaction models have been conducted. In the context of physically situated agent, Matsusaka et al. pioneered multiparty conversational robots (Matsusaka, Tsuyoshi, and Kobayashi 2003). They developed a conversational robot fulfilling requirements to participate in a group conversation, including understanding status of participation, and ways of transferring messages to other participants. Bohus et al. pioneered the Open World Dialogue, “centering on building and studying systems that can engage in conversation in an open-world context, where multiple people with different needs, goals, and long-term plans may enter, interact, and leave an environment” (Bohus and Horvitz 2009). Foster et al. presented another similar interactive kiosk system that handles situated, open-world, multimodal dialogue in scenarios, namely the JAMES, Joint Action for Multimodal Embodied Social Systems (Foster et al. 2012). It extends the Bohus’s framework by adding physical embodiment, which has been shown to have a large effect on social interaction.

∗Yoichi Matsuyama is also a post doctoral research fellow at the ArticuLab, Human-Computer Interaction Institute, Carnegie Mellon University.

Copyright © 2015, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.
to the other participants. In this paper, we propose a procedural facilitation process framework to harmonize a four-participant conversational situation. The situations and procedures are modeled and optimized as a partially observable Markov decision process (POMDP), which is suitable for real-world sequential decision processes, including dialogue systems (Williams and Young 2007).

The remainder of this paper is organized as follows. We begin by reviewing facilitation frameworks in small groups and describing procedures for maintaining small groups. Then we give an overview of the architecture of our proposed system. We then discuss three experiments conducted to verify the efficacy of the small group maintenance procedures and the performance of POMDP. Finally, we summarize and conclude this study.

Facilitation Framework

In this section, in order to organize the facilitation framework, at first, we review related works of facilitation models in small groups, specifically functional roles of group members that have been defined to analyze facilitation processes. Then we review engagement models, and we propose the harmony model.

Small Group Maintenance

Benne et al. analyzed functional roles in small groups to understand the activities of individuals in small groups (Benne and Sheats 1948). They categorized functional roles in small groups into three classes: Group task roles\(^1\), Group building and maintenance roles\(^2\), and Individual roles\(^3\). The Group task roles are defined as “related to the task which the group is deciding to undertake or has undertaken,” whose roles address concerns about the facilitation and coordination activities for task accomplishment. The Group building and maintenance roles are defined as “oriented toward the functioning of the group as a group,” which contribute to social structures and interpersonal relations. Finally, the Individual roles are directed toward the individual satisfaction of each participant’s individual needs. They deal with individual goals that are not relevant either to the group task or to group maintenance. In this paper, we specifically focus on Benne’s Group building and maintenance roles. To coordinate situations, we assume a facilitator must take the following procedural steps as group building and maintenance roles.

1. Observation: To be aware of both the presence of dominant participants leading the current conversation and the status of a left-behind participant.

2. Initiative Acquisition: To obtain an initiative to control the situation and wait for approval from the others, either explicitly or implicitly.

3. Floor and Topic Shift: To give the floor to a suitable participant (sometimes by initiating a new topic).

We will formalize the procedure in more depth below.

Engagement Density

In order to formalize procedural steps obtaining an initiative controlling a situation, we begin by extending the participation structure model in multiparty conversations. The participation structure model was presented by Clark (Clark 1996), drawing on Goffman’s work (Goffman 1981). In this model, each participant is assigned a participation role considered by the current speaker, where speaker, addressee, and side-participant are “ratified participants.” Ratified participants include the speaker and addressees, as well as a side-participant who is taking part in the conversation but is not currently being addressed. All other listeners, who we refer to as over-hearers, have no rights or responsibilities within the structure. Over-hearers come in two main types. Bystanders are those who are openly present but not part of the conversation. Eavesdroppers are those who listen in without the speaker’s awareness. The speaker must pay close attention to these distinctions when speaking. For example, the speaker must distinguish addressee from side-participants. When the speaker asks an addressee a question, the speaker must make sure that it is the addressee who is intended to answer the question, and not side-participants. However, the speaker must also ensure that the side-participant understands the question directed at the addressee. In addition, the speaker must consider the over-hearers. However, because the over-hearers have no rights or responsibilities in the current conversation, the speaker can treat them as he pleases.

In this paper, we extend Clark’s model with the concept of engagement. In terms of engagement among conversational participants, Sidner et al. dealt with engagement in multimodal ways, including eye gaze. They defined engagement as “the process by which two (or more) participants establish, maintain and end their perceived connection during interactions they jointly undertake” (Sidner et al. 2004). This process includes: (1) initial contact, (2) negotiating a collaboration, (3) checking that other is still taking part in the interaction, (4) evaluating whether to stay involved, and (5) deciding when to end the connection. Based on these previous studies, we define engagement as the process establishing connections among participants using dialogue actions so that they can represent their own positions properly.

In Figure 1, suppose participant C has been assigned as a side-participant who has not engaged with other participants for a significant time. Participant C’s amount of communication traffic with the other participants is significantly less than that of the others. Here, we define “engagement density,” which represents the amount of communication traffic. As a relevant measurement of engagement density, Katzenmaier et al. produced a measure of “utterance density,” which takes the ratio of speech to non-speech be-

---


2. Group building and maintenance roles: Compromiser, Harmonizer, Standard setter, Gatekeeper and expediter, Encourager, Observer and commentator, and Follower.

haviour per utterance (“a speech activity per a certain unit of time by dividing each utterance duration by the sum of previous and following pause durations”) (Campbell and Scherer 2010). While the utterance density directly depends on speech activities, the engagement density is a measurement of amount of communication between interlocutors. Therefore, even if a participant’s utterance density is high, it does not mean the engagement density is high. Jokinen et al. also mentioned that sometimes one of the participants might be less active in turn-taking (engagement) even if the speaking activity in the conversation as a whole is large (Jokinen 2011). Three-participant conversations are likely to produce a difference of density. We define a “harmonized” participant as a participant with high engagement density, and an “un-harmonized” participant as a participant with low engagement density. Consequently, speaker and addressee are always assigned as harmonized participants, and side-participants can be divided into two types in terms of engagement density: harmonized side-participant and un-harmonized side-participant. Although all side-participants are ratified, an un-harmonized side-participant, who is only recognized by the speaker, can sometimes emerge in four-participant situations.

Procedures for Engagement Density Control

In order that a facilitator is transferred an initiative by the current speaker, the facilitator must take procedural steps. First, the facilitator must participate in the current dominant conversation the speaker is leading, try to be “harmonized” to claim an initiative, and then wait for either explicit or implicit approval from the speaker. Let us take the example shown in Figure 1. In the figure, participants A and B are primarily leading the current conversation. Participant C cannot get the floor to speak, and so the robot desires to give the floor to C. If the robot who is an “un-harmonized” participant speaks to C directly, without being aware of A and B, the conversation might be broken, or separated into two (A-B and C-robot), at best. In order not to break the situation, the robot should participate in the dominant conversation between A and B first, and set the stage such that the robot is approved to initiate the next situation as “harmonized” participant. In this paper, we assume that every person participating in a dominant conversation is at “harmonized” state (participant A, B in Figure 1), and the other is at “un-harmonized” state (participant C and a robot). After participating in the dominant conversation between A and B, the robot is approved as a “harmonized participant” to initiate the conversation.

In terms of the way of controlling engagement, Whittaker et al. analyzed two-participant dialogues to investigate the mechanism how each control was signaled by speakers and how it affects discourse structure, including the lower control level, topic level and global organization level (Whittaker and Stenton 1988). For the control level, they found that three types of utterances (prompts, repetitions and summaries) were consistently used to signal. For the topic level, they found that interruptions introduce a new topic. And the global organization is organized also by topic initiation. This study argued that not only signal utterances but also topic shifting_INITIALIZATION plays an important role for engagement control. On the basis of these discussions above, we define the following constraints for both harmonized and un-harmonized participants when they address a next speaker and shift current topics:

1. **Constraint of addressing:**
   An un-harmonized participant must not address the other un-harmonized participants directly.

2. **Constraint of topic shifting:**
   An harmonized participant must not shift the current topic when he/she addresses the other un-harmonized participants.

For examples, while a harmonized participant (speaker, addressee and harmonized side-participant) can address an both harmonized (addressee and harmonized side-participant) and un-harmonized (un-harmonized side-participant) participants, an un-harmonized participant can not address another un-harmonized participant. In the following sections, we describe a computational model that has the group maintenance functions discussed above.

Timing of Initializing a Procedure

In order to detect timing of initializing a procedure, a facilitator should care about a unit of consecutive sequence to avoid to break a current conversation. An adjacency pair is a minimal unit of conversational sequence organization (Sacks 1973), therefore it might be reasonable to employ here. An adjacency pair is characterized by certain features (Schegloff 2007): a) composed of two turns, b) by different speakers, c) adjacently placed, d) these two turns are relatively ordered;

So, which timing can be candidates for a facilitator to initiate procedures? As a facilitator might produce economically short steps of procedures to help a left behind participant, in this paper, we assume every second or third part might be the candidates to initiate. Figure 3 shows transition of harmony state, which describes how a facilitator makes himself/herself harmonized and takes an initiative to control a situation, by employing a concept of adjacency pairs. We assume that an Un-Harmonized participant needs to be approved by a speaker’s second pair part to be harmonized. In the following sections, we will describe a computational model of the procedural process discussed above.

Architecture for Facilitation Robots

We propose a computational architecture for multiparty conversation facilitation robots, namely the SCHEMA Framework. The SCHEMA Framework mainly consists of the following processes: the Perception Process the Procedural Production Process the Language Generation Process. The Perception Process process interprets situations based on visual and auditory information. This process includes Adjacency Recognition, Participation Recognition, Topic Recognition and Question Analysis. The Procedural Production Process produces procedural actions to manage a group, referring Goal Management Module. This module is modelled as a reinforcement learning framework (partially ob-
servable Markov decision process (POMDP)). The Perception Process and the Procedural Production Process will be described in the next section. The Language Generation Process has Content Planning, Sentence Planning (Answer and Question Generation), Behavior Planning and Realization processes. These processes will be described in latter section.

Procedure Optimization using POMDP

To optimize the procedures discussed above, we model the task as a partially observable Markov decision process (POMDP) (Williams and Young 2007). Formally, a POMDP is defined as a tuple \( \beta = \{ S, A, T, R, O, Z, \eta, b_0 \} \), where \( S \) is a set of states describing the agent’s world, \( A \) is a set of actions that the agent may take, \( T \) defines a transition probability \( P(s'|s, a) \), \( R \) defines the expected reward \( r(s, a) \), \( O \) is a set of observations the agent can receive about the world, and \( Z \) defines an observation probability, \( P(o'|s', a) \), \( \eta \) is a geometric discount factor \( 0 < \eta < 1 \), and \( b_0 \) is an initial belief state \( b_0(s) \). At each time-step, the belief state distribution \( b \) is updated as follows:

\[
b'(s') = \eta \cdot P(o'|s', a) \sum_s P(s'|s, a)b(s) \tag{1}
\]

In this paper, we assume \( S \) can be factored into three components: the participants’ engagement states \( S_e \), the participants’ motivation states \( S_m \), and the participants’ actions \( A_p \). Hence, the factored POMDP state \( S \) is defined as

\[
s = (s_e, s_m, a_p) \tag{2}
\]

and the belief state \( b \) becomes

\[
b = b(s_e, s_m, a_p) \tag{3}
\]

To compute the transition function and observation function, a few intuitive assumptions are made:

\[
P(s'|s, a) = P(s'_e, s'_m, a'_p|s_e, s_m, a_p, a_o) = P(s'_e|s_e, s_m, a_p, a_o), \tag{4}
\]

\[
P(s'_m|s'_e, s_e, s_m, a_p, a_o), \]

\[
P(a'_o|s'_m, s'_e, s_e, s_m, a_p, a_o)
\]

The first term in (4), which we call the harmony model \( T_s_h \), indicates how participants harmonize in the current dominant conversation at each time-step. We assume that the participants’ harmony state at each time-step depends only on the previous harmony state, the participants’ action, and the system action. The transition probability can be described as follows:

\[
T_s_h = P(s'_h|s_h, a_p, a_o) \tag{5}
\]

In this paper, the probabilities of (5) were handcrafted, based on the consideration in Section and our experiences. When the harmony state is the Un-Harmonized state and the robot is asked by a current speaker, the state should be changed to the Pre-Harmonized state, where the robot is awaiting the speaker’s approval for the Harmonized state. We assume that any dialogue acts from the speaker addressing the robot in the Pre-Harmonized are approvals. Otherwise, the state will be back to the Un-Harmonized. The Harmonized state gradually goes down to the Un-Harmonized state in time-steps unless the robot selects any dialogue acts.

We call the second term the participants’ motivation model \( T_s_m \), which indicates how an Un-Harmonized participant has the motivation to take the floor at each time-step. This state implies that an unharmonized participant has a motivation to speak on the current topic. Thus, this state affects decision-making about topic shift. We assume that a participant’s motivation at each time-step depends only on the previous system action. The motivations are defined as an unengaged participant’s ID and a binary (true/false) variable.
We call the third term the gradually falling down to “participant who claimed an initiative. (4) An “ was approved by the speaker’s second pair part addressed to the participant who claimed an initiative. (3) A claim was implicitly approved by the speaker’s second pair part. (2) A claim was assigned as a side-participant.

Figure 3: Transition of harmony states. (1) A participant claims an initiative with a first pair part, against a current speaker who is leading the current dominant conversation, waiting for either explicit or implicit approval by the speaker’s second pair part. (2) A claim was declined by the speaker either explicitly or implicitly. (3) A claim was implicitly approved by the speaker’s second pair part. (4) An “Harmonized” state is gradually falling down to “Un-Harmonized” while a participant is assigned as a side-participant.

Table 1: Robot’s harmony states $s_h$

<table>
<thead>
<tr>
<th>Harmony states</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Un-Harmonized</td>
<td>The robot is not harmonized with the current conversation.</td>
</tr>
<tr>
<td>Pre-Harmonized</td>
<td>The robot is waiting for approval to harmonize with the current conversation.</td>
</tr>
<tr>
<td>Harmonized</td>
<td>The robot is harmonizing with the current conversation.</td>
</tr>
</tbody>
</table>

Table 2: Un-Harmonized participant’s motivation states $S_m$

<table>
<thead>
<tr>
<th>Motivation states</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motivated</td>
<td>The participant who is left behind has a motivation to speak on the current topic (interested in the current topic).</td>
</tr>
<tr>
<td>Not-Motivated</td>
<td>The participant who is left behind does not have any motivation to speak (not interested in the current topic).</td>
</tr>
<tr>
<td>None</td>
<td>Nobody is left behind.</td>
</tr>
</tbody>
</table>

\[
b'(s'_m, a'_p) = \eta \cdot P(a'|s'_m, a_p, a_s) P(s'_m|a_s) \sum_{a_p} P(a'_p|s'_h, a_p, a_s) \sum_{s_h} P(s'_h|s_h, a_p, a_s) b(s'_m, a_p) \tag{9}
\]

On the basis of the consideration of the constraints in Section, the reward measure includes components for both the appropriateness and inappropriateness of the robot’s behaviors. The demo video of the proposed system is on Youtube.4

**Situation Interpretation**

**Participation Role Recognition**

In this paper, we employ the following assumptions for role classification in a four-participant situation.

1. One *speaker* always exists in one group at each time-step.
2. One *addressee* who is addressed by the *speaker* always exists at each time-step.
3. A *side-participants* is a participant who is not assigned neither *speaker* nor *addressee*.

As we defined in the privies section, *side-participants* can be divided into two types: *harmonized side-participant* and *un-harmonized side-participant*. The recognition process consists of distinctive three sub-processes: speaker classification, addressee classification, and harmonized/un-harmonized side-participant recognition. Many research mentioned that acoustic and visual cues, such as gaze direction, face direction, head pose and acoustic information are reliable cues for addressing in multiparty human-human and human-robot interactions (Katzenmaier, Stiefelhagen, and Schultz 2004) (Jovanović, Nijholt, and others 2006) (Fujie, Miyake, and Kobayashi 2006) (Johansson, Skantze, and Gustafson 2013). In this paper, the speaker classification is based on the results of face direction classification and VAD.

4https://www.youtube.com/watch?v=oanbOmNida0&list=UU28s8YBlgfSRfEdfdA0tw
The address classification is based on the result of speaker classification, as well as each participant’s face direction and VAD. The face directions are captured by depth-RGB cameras (Microsoft Kinect). The best results of classification using Naive Bayes for speaker and address classification were 79.4% and 70.9%, respectively.

In the final process, another participant, who should be assigned to a side-participant according to our definition above, is estimated whether he/she is harmonized or unharmonized. In the scenario shown in Figure 3, participant C may not be able to take the floor for a while. We assume the situation probably resolves itself when the current topic is shifted. Hence, we define the depth of side-participation as follows:

\[ \text{Depth}_{\text{SPT}} = \frac{\text{Duration}_{\text{SPT}}}{\text{Duration}_{\text{topic}}} \]  

(10)

\[ \text{Harmonized}_i = \begin{cases} 1 & \text{if } \text{Depth}_{\text{SPT}_i} > \text{Threshold} \\ 0 & \text{otherwise} \end{cases} \]  

(11)

where the suffix \( i \) represents a participant’s ID.

**Motivation Estimation** The motivation estimation manages only an un-harmonized participant’s motivation to take a floor on the current topic. Thus, this state affects decision making about topic maintenance. We define motivation as the un-harmonized participant’s ID and a binary (true/false) variable, which is heuristically calculated as follows:

\[ \text{Motivation}_i = \begin{cases} 1 & \text{if } \text{MotivationAmount}_i > \text{Threshold} \\ 0 & \text{otherwise} \end{cases} \]  

(12)

In our previous experiment, we analyzed how a conversational robot’s existence and its actions can affect users’ emotional impressions in group game situations, using video analysis, SD (Semantic Differential) method and free-form questionnaires. The result of SD method indicates that subjects feel more pleased, and the results of free-form questionnaires showed many participants were motivated to participate in the game, with participation and active actions of a robot. These psychological results correlate with utterance frequency and smiling duration ratio, calculated by annotated data (Matsuyama et al. 2010). Also, according to our observation and discussions of the experiments, even if participant’s utterances are not observed frequently, participants motivated to participate are likely to nod frequently, as reacted to a speaker’s utterances. Therefore, we assume the amount of motivation of a participant can be calculated by a heuristic linear function of speech, smiling and nodding activities during duration of a certain topic, as follows:

\[
\text{MotivationAmount}_i = \int_{t_{\text{start}}^i}^{t_{\text{end}}^i} (\alpha f_{\text{speech},i}(t) + \beta f_{\text{smile},i}(t) + \gamma f_{\text{nod},i}(t)) dt
\]

(13)

where \( t \) represents a current time, \( t_{\text{start}} \) and \( t_{\text{start}} \) represent start and end times of a continuum topic, respectively. \( \alpha, \beta \) and \( \gamma \) are arbitrary coefficients. The speech activities are calculated using results of VAD. The smiling and nodding activities are calculated by smiling detection and nodding detection modules, using Microsoft Kinect’s Face Tracking SDK.

**Adjacency Pairs Recognition** In this paper, adjacency pairs are recognized by the results of participation role recognition and speech recognition. Each time the system detects an endpoint of speech from the automatic speech recognition module, it classifies each utterance into one of the six categories shown in Table 3 (1st, 2nd, 3rd) × {toRobot, ntoRobot}). In this paper, adjacency pairs are recognized by the linear-chain conditional random fields (CRF), using results of speech recognition.

For learning and evaluation, we recorded conversational data where 3 participants are assigned to each group and talked for 10 minutes. We had totally 7 groups (70 minutes with 21 participant). They were instructed that they would talk about movies within movie-related 100 topics we defined beforehand. We used 6 groups for learning, 1 group for evaluation. After we transcribed the recorded conversations, each utterance separated manually by an experimenter. Then each of them is analyzed by a Japanese language morphological analyzer\(^3\). The analyzer allows the part of speech to be further sub-classified, namely the subparts of speech. Based on the analyzed results, we coded each morpheme with an extended BIO encoding scheme. Using the BIO, each word is tagged as either (B)eginning an entity, being (I)n an entity, or being (O)utside of an entity. In this case, we extended it with adjacency pairs: a beginning of a first pair part is coded as “B-1”, and subsequent words are coded as “I-2”. The same rule is applied for both second and third parts (“B-2” or “I-3” for second parts, “B-3” or “I-3” for third parts). As for

\(^3\)http://nlp.ist.i.kyoto-u.ac.jp/index.php/JUMAN
the successfulness of the coding, the inter-rater agreement using Cohen’s kappa (Fleiss, Levin, and Paik 2013) indicated a substantial result between the two raters ($\kappa = 0.75$). The classification accuracy for each word was 73.5%. And a result of the last word will be the final result of the adjacency pair.

Language Generation

**Topic Management** In this paper, we define a sequence of topic words as a conversational context. Each system utterance is hooked to one of the topics. For example, the sentence “Audrey is beautiful, isn’t she?” is assumed to belong to the topic “Audrey Hepburn.” In our experimental system, we prepared 100 topic words for each domain. The topics in the *movies* domain include genres, titles, directors, and actors.

The topic estimation procedure uses the following three processes: Japanese language morphological analysis, important words filtering, and classification. After an ASR or text input is processed by Japanese language morphological analysis, only nouns are extracted. Then, the important nouns in each topic are extracted. In terms of degrees of importance, we use the term frequency-inverse document frequency (TF-IDF) score, which is often used as a weighting factor in information retrieval and text mining. We collected the top 64 web sites as 64 separate documents for each topic word using Google web search. In the classification process, we used the linear-chain conditional random fields (CRF) technique. As an evaluation experiment, we evaluated the accuracy of 10-topic classification. We recorded three-minute conversations with two participants in which they were instructed to talk within 10 topics in the animation film domain. We conducted a total of 25 sessions, 20 of which were used for learning data, and five used for test data. One experimenter annotated each word as correct answers. The result for the accuracy rate (number of correct answers / total number estimated) was 88.2% under a word error rate for ASR of 0%, and 64.7% under a word error rate for ASR of 20%.

**Question Generation** The Question Generation Module has two main functions: giving someone the floor and collecting the user model. The user model is preferred for topic maintenance. We define that a user’s interests in a certain topic are organized by experiences and preferences. The system extracts this information in the following ways.

1. User’s answer to the system’s question.
   The system directly asks a user his/her experiences and preferences about a certain topic.
2. User’s motivation (interests) for each topic.
   When a topic transition occurs, the system obtains each user’s preference, which is calculated as the sum of their motivation during the topic.

   A preferred new topic is determined using cosine similarity of TF-IDF scores. The topic scores ($\text{TopicScore}_i$) of all topics are calculated on the basis of the cosine similarities of the current topic ($\text{CurrentTopic}$), a user’s topic preferences of all topics ($\text{PreferenceTopic}_i$), and experiences ($\text{ExperienceTopic}_i$) between the $\text{CurrentTopic}$ and each $\text{Topic}$.

   \[
   \text{TopicScore}_i = \alpha \cos(\text{Topic}_i \cdot \text{CurrentTopic}) + \beta \left( \sum_m \cos(\text{Topic}_i \cdot \text{PreferenceTopic}_m) \right) + \gamma \left( \sum_n \cos(\text{Topic}_i \cdot \text{ExperienceTopic}_n) \right)
   \]

where $\alpha > \beta > \gamma$. According to the $\text{TopicScore}$, the system can shift a topic to another that is close to a left-behind participant’s interest.

**Answer Generation** Based on the results of the Question Analysis process, answers are classified into two types: Factoid type answers and Non-factoid type answers (opinions). Factoid answers are generated from a structured database. In this research, we use Semantic Web technologies. After analyzing a question, it is interpreted as a SPARQL query, a resource description framework (RDF) format query language to search RDF databases. We use DBpedia as an RDF database. The opinion (non-factoid type answers) generation process refers opinion data automatically collected from a large amount of reviews in the Web. The opinion generation consists of four process: document collection, opinion extraction, sentence style conversion, and sentence ranking. As an example task, we collected review documents from the Yahoo! Japan Movie site. For further explanations of the mechanisms of the Answer Generator, see (Matsuyama et al. 2015). The demo video of the proposed system is on Youtube.

### Experiments

**Appropriateness of Procedures**

In order to evaluate the efficiency of our proposal procedure, we designed the following two experiments. The experiment 1 evaluates the appropriateness and feeling of groupness as results of our proposed procedures with 35 subjects. The experiment 2 evaluates the appropriateness of timing of initiating procedures with 32 subjects.

In the first and second experiments, we prepared videos of four-participant conversational situations (Human person A, B, C, and a robot), where a facilitation robot initiates procedures, or naively approaches the un-harmonized participant C without procedural steps. The spatial arrangement was the same as the previous sections. Each subject was requested to watch videos from a third party. The results in the experiment 1 indicate that the usage of procedures to obtain initiative before approaching an unharmonized participant showed evidence of acceptability and feeling of groupness. Regarding timing issue in the experiment 2, initiating the procedures just after the second or third adjacency pair part is considered more appropriate than that after the first pairs.

We also conducted simulation experiments comparing POMDP and MDP based systems to evaluate how a robot

---

6http://ja.dbpedia.org/
7https://www.youtube.com/watch?v=oaNhOmNida0&list=UU28s8YBlqfISRJfOEIdfOAotw
could properly approach an unharmonized participant under recognition error conditions. As the results, while the MDP-based system was more sensitive to observation errors, the POMDP-based system could cope with errors.

Effectiveness of Opinion Sentence Generation

We conducted two experiments to evaluate grammatical acceptability and effectiveness of sentence generation algorithms. In the first experiment, five subjects were requested to read and rate each sentence extracted and ranked. As the result, the automatically extracted and ranked opinions we proposed got 78.8%, 75.7% and 73.6% of acceptability in Short, Standard and Diverse algorithms, respectively. In the second experiment, 38 subjects were requested to watch four types of pre-recorded videos (Short, Standard, Diverse and Randomly-Mixed), where a robot and a human participant were talking each other, and requested to answer questionnaires after watching. The results implies that the Diverse utterances attracted users’ interests and made them willing to talk, making the robot more favorable.

Conclusions

In this paper, we presented a computational model for facilitation process in a small group. The model included regulating socially imbalanced situations with procedural decision process controlling engagement density to harmonize a group as a group, and language generation process associated with user models. These situations and procedures were modeled and optimized as a POMDP. As the result of two user experiments, usages of procedures obtaining initiatives showed evidences of acceptability as a participants behaviors, and feeling of groupness. As for timings, initiating the procedures just after the second or third adjacency pair parts is felt more appropriate than the first pairs by participants.

The future work include considering extensions of POMDP model for task goal management, while we discussed mainly aspects of group maintenance for facilitation in this paper. We are also considering extending the participation role recognition module adjacency pair recognition and motivation estimation, using gaze direction information, mentioned its importance in related work.

Acknowledgement

This research was supported by the Grant-in-Aid for scientific research WAKATE-B (23700239) and JSPS Strategic Young Researcher Overseas Visits Program for Accelerating Brain Circulation. TOSHIBA corporation provided the speech synthesizer customized for our spoken dialogue system. We also thank Shinya Fujie and Iwao Akiba for their substantial contributions.

References


