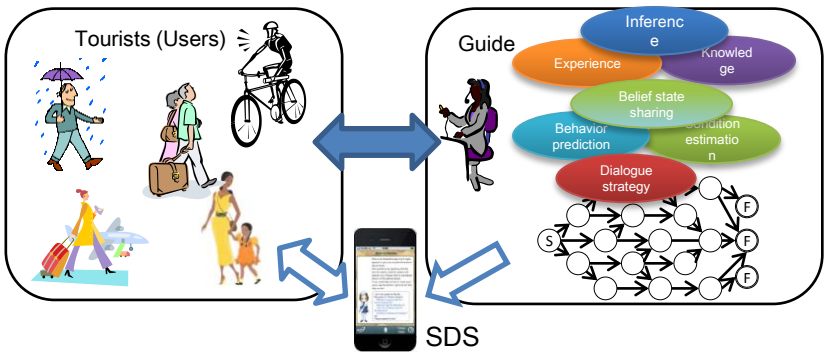
What are they achieving through the conversation?
- modeling guide-tourist dialogues by extended grounding networks

Etsuo Mizukami and Hideki Kashioka
National Institute of Information and Communications Technology
mailto:etsuo.mizukami@nict.go.jp

Introduction

To engage conversations, especially those which have certain goals or tasks, the participants have to share many things to achieve collaboration. In each local interaction, the achievement of *grounding* (Clark and Schaefer, 1989; Clark and Brennan, 1991) about sharing each utterances meanings would be needed. By accumulating these grounding process, the final goal of the conversation would be accomplished. It is needed that both an *evidence* which the addressee would understand the speaker's utterance is exhibited and they mutually share it. To develop a collaborative dialogue system, it is important that the system is designed to consider this grounding process. In this study, we propose a dialogue modeling method to deal with the grounding process between participants for goal-oriented dialogues.



Ideas based on

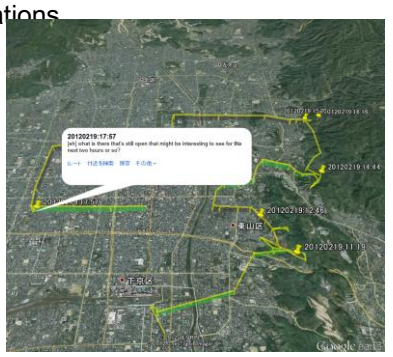
Clark and Schaefer (1989) formulated a *contribution* model in order to model the grounding process. The *contribution* consists of two phase of *presentation* and *acceptance*, e.g., an initiation 'hello' and a response 'hello'. Traum (1996) extended this idea to a computational model of grounding as a transition network. While the conversation has a *discourse purpose* as goal and the hierarchical and multiple *discourse segment purposes* to be shared between one and another in each *discourse segments* (Grosz and Sidner, 1986). By merging these ideas, we would like to propose the idea of *contribution topics* and *extended grounding networks*.

Annotation scheme

Contribution topics -- a unit of achievement corresponding to a discourse segment which has a certain preposition to be shared with collaborators for a certain dialogue period. It also has a hierarchical structure where a contribution topic of an higher levels contain the topics from middle and lower levels.
Grounding Acts -- based on Traum's seven grounding acts: *initiate, continue, ack, reqrepair, reqack, repair* and *cancel,* they are labeled to each utterance in the bottom level of contribution topics.
Grounding states -- strength of grounding is expressed by four kinds of states: ungrounded (-), half-grounded (!), mid grounded ( ) and mutual grounded (m).

Objective Data

We applied the annotation scheme to the data recorded from cell-phones; conversations held between a guide and ten tourists in Japanese. We had the tourists travel around Kyoto City for a full-day, and had them to call the guide several times during their trips. Thirteen of them (now, we have 98) were labeled and analyzed.



An example of labeled data

Each lowest level contribution topic is initiated from start state S to final state F (= m) through each collaborator's mid-grounded states g (guide's state) or u (user's state) or other (half-grounded, ungrounded) states. If a lower level of contribution topic is completed, the state of the upper level can be transitioned to the next state.

R	Utterance	U-ID	d=1	State	d = 2 (1)	State	d = 3 (1)	State	d = 2 (2)	State	d = 3	State	d = 2 (3)	State	d = 3	State
				S		S		S								
U	[Ring]	0	<opening>		<channel>		initiate	u				(S)				
G	Hello, this is Hiromi speaking.	10					ack	g	<identify>		initiate	g				
U	This is Tanaka speaking.	20			</channel>	1	ack	m			ack/initiate	u				
G	How are you?	30							</identify>	2	ack	m/S	<social>		initiate	g
U	Excuse me for calling many times.	40													ack/initiate	u
G	No problem.	50	</opening>	1									</social>	F/S	ack	m/S
U	So, I'm in the grounds of Kiyomizu temple.	60	<request>		<situation>		initiate	u								
G	Yes.	70			</situation>	S	ack	m/S								
U	I have just drawn a paper fortune,	80			<und:req>		initiate	u								
G	uh-huh	90					ack	g								
U	and I forgot whether I should bind a bad one or a good one with a tree branch.	100					continue	u								
G	Please bind bad one.	110			</und:req>	1	ack	m/S	<answer>		initiate	g				
U	Should I bind bad one, and take good one along?	120									reqack	u				
G	Yes, you should.	130									ack	g				
U	I got it.	140	</request>	2					</answer>	F/S	ack	m/S				
U	Thank you very much.	150	<closing>		<social>		initiate	u								
G	Was it good result?	160			</social>	1	ack	m/S	<addition>		initiate	g				
U	Yes, it was good.	170									ack/initiate	u				
G	Please keep it.	180									ack/initiate	g				
U	Yes, thanks.	190							</addition>	2	ack	m/S	<social>		initiate	u
G	Sure, anytime.	200											</social>	3	ack/	m/S
U	Goodbye.	210											<social>		initiate	u
G	Goodbye.	220	</closing>	F									</social>	F	ack/	m

Results

Extracted Contribution Topics (CTs) by levels of network

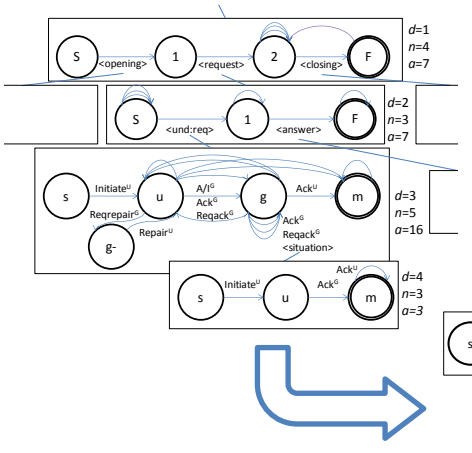
Top level (depth d=1) *Input symbols of this level consist of only CTs
- <opening>
- <request>
- <closing>
- <proactive> *guide's unprompted proposal
- <appointment> *scheduling continuous calls

Middle level (d=2)
- <und:req> *understanding user's request
- <answer>

Bottom level (d≧2) *only Grounding Acts
- <channel>
- <identify>
- <social>
- <pre:req> *preface of request
- <sub:req> *user's supplementary request
- <probe> * guide's follow-up question

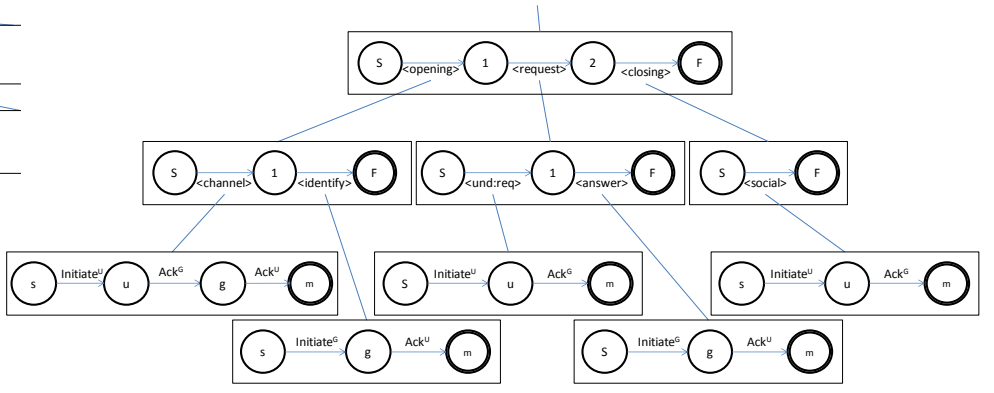
Multi level (d=any)
- <situation> *confirming the situation of the user
- <goal> *reference to goal of the trip
- <addition> *additional chat

An example of network



Basic network model of Guide-Tourist dialogue

From the communality of all networks, the most simplest (or indispensable) network can be conducted as the basic model of the guide-tourist dialogues

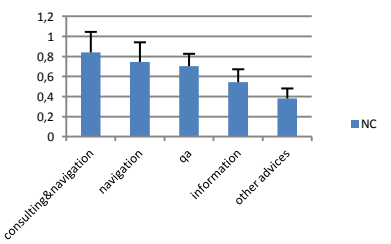


Discussion

The actual human-to-human dialogues include many redundant exchanges. In the extended grounding networks, those are expressed as additional transitions to the basic network. In a sense, such redundancy can be regarded as naturalness of human-to-human dialogue, but some part of them should (and might) be reduced in some way (by using sensor information like GPS, camera, etc). It may be estimated with the network complexity (NC).

$$NC = \sum_{i=1}^{k} \frac{1}{d_i} \log\{(n_i - nbase_i + 1)(a_i - abase_i + 1)\}$$

NCs of <request> by request types



Future works

Since the model has difficulty to deal with parallel topics proceeding simultaneously at the same level (e.g. <opening> topic), we must extend further our framework. Besides we have to undertake
- automatic extraction of basic networks
  automatic labeling from DA or other labels
  implementing the model as dialogue strategy