

Collection and Analysis of Travel Agency Task Dialogues with Age-Diverse Speakers

Michimasa Inaba, Yuya Chiba, Ryuichiro Higashinaka, Kazunori Komatani, Yusuke Miyao, Takayuki Nagai

Introduction

- When individuals communicate with others, they use different vocabulary, speaking speed, facial expressions, and body language depending on the people they communicate with.
- We focus on the **speaker's age** as a factor that dramatically affects the change in communication.
- We collected a **multimodal dialogue corpus with a wide range of speaker ages**.
- Dialogue task based on a tourism consultation at a travel agency.



Travel Agency Task

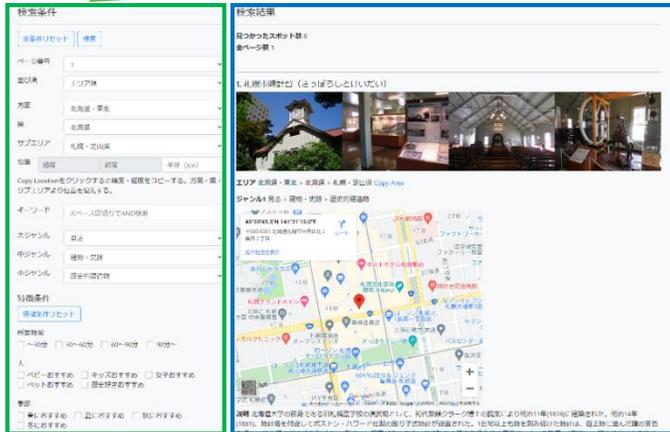
Dialogue Task

- Dialogue between two speakers
- One playing the role of an operator and the other playing the role of a customer, simulating a tourist consultation at a travel agency.
- The operator recommends tourist spots by using information obtained from **the tourist information retrieval system**

Tourist Information Retrieval System

- Developed using the Rurubu data API provided by JTB Publishing Corporation.
- Rurubu data contain approximately 45,000 Japanese sightseeing spots.

The operator can specify search queries such as region and area, free keyword search, genre-based search and budget.



Photos, maps, business hours, transportation access, etc. of tourist spots

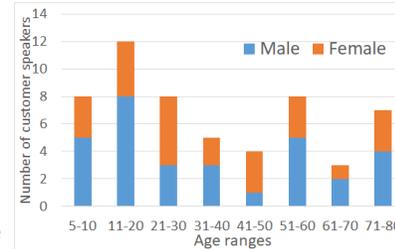
Data Collection and Annotation

Recording

- Recorded using Zoom
- 330 dialogues (20 minutes each)

Speakers

- Operators : 5 people
- Customers: 55 people
 - Minors (7-17) 20 people
 - Adults (20-60) 25 people
 - Older Adults (65-72) 15 people



Transcription and Annotation

- The collected dialogues were manually transcribed and annotated with dialogue act (DA) tags via crowdsourcing.
- A subset of the ISO 24617-2 annotation scheme (the first edition) [Bunt+ 17] was used as the DA tag set (44 tags)
 - Cohen's κ was 0.632 (Good agreement)

Corpus Statistics

Dialogues	330
Duration	115 hour 48 minutes
Utterances (turns)	111,771
DA Tags	246,316

Dialogue Examples

Well, [Stalling] there are a lot of places around the market where you can eat sushi. [Inform]

Yes. [AutoPositive]

What kind of sushi do you want to eat? [PropositionalQuestion]

Uh, [Stalling] a sushi restaurant has a lots of salmon roe, like a bowl of salmon roe. [Answer]

Is it hard for you to walk? [PropositionalQuestion]

Yeah, [AutoPositive] I used to love climbing mountains when I was younger, but my knees are getting worse with age. [Answer]

Ah, [Stalling] there are various, [CheckQuestion] uh, [Stalling] types of houses with kabuki roofs, what do you think? And there is also a place like a museum, [CheckQuestion] uh, [Stalling] it seems to exhibit these old houses. [CheckQuestion]



Minor customer



Operator



Older Customer



Operator

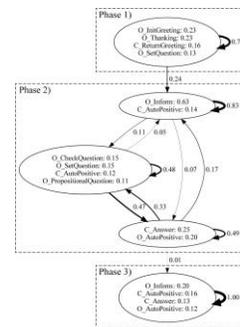
Dialogue Analysis

Analyzed how the customer's behavior differs between age groups

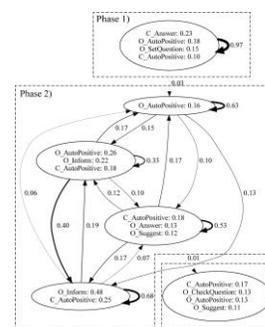
Analysis of DA sequence

- Modeled by the hidden Markov model (HMM)
 - Initial and final phases:** Single state with a self-loop
 - Middle phase:** Ergodic HMM

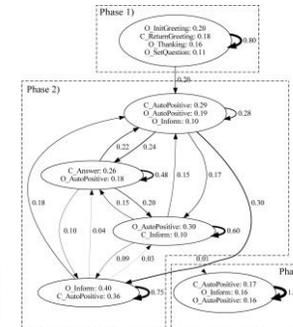
Minor customers



Adult customers



Older customers



- Minor customers:** Simple structure with operator gives information and customer responds to it.
- Adult customers:** One of features is a state with a high output probability of operator's suggestion \Rightarrow operators provide customers constructive guidance
- Older customers:** There is a state in which customer gives information, and operator responds to it \Rightarrow older customers tend to talk about their motivations and wishes for trip

Analysis of facial expression

- Compared average value of AUs (ANOVA + multiple comparison test)
- Operator expressed a smile more frequently to minor customers

Results of multiple comparison tests for operator's facial expression

AU	Comparison	Diff.	p-value
AU06	Minor-Adult	0.382	< 0.001***
	Minor-Older	0.350	0.011**
	Adult-Older	-0.031	1.000
AU12	Minor-Adult	0.507	< 0.001***
	Minor-Older	0.423	0.023*
	Adult-Older	-0.084	1.000

*p < 0.05, **p < 0.01, ***p < 0.001

The system should therefore adapt not only verbal but also non-verbal dialogue behavior to the user's age to provide natural travel guidance.

Conclusions

- We constructed and analyzed multimodal dialogue corpus with a wide range of speakers' ages from children to the elderly.
- We analyzed the sequences of dialogue acts and the facial expressions focusing on the age of the speakers.
- Working on making the corpus accessible to the public.

The data collection procedure has been approved by the ethics committee for experiments on human subjects, the University of Electro-Communications (No. 19061(2)).