# Talking with AIs about views from a vehicle

Abstract

This is a study of a partner dialogue system for mobile vehicles that uses the ever-changing, real-time view from the car as a topic of conversation. This system uses a deep-learning based dialogue model using the largest scale of Japanese dialogue data developed by NTT to realize natural dialogue. This system integrates scenery images from vehicle and spots around the car's location to talk about scenery around the vehicle. By sequentially incorporating information about the area around the vehicle's location, we are realizing a new experience of driving while enjoying the pleasure of sharing the "now" with a knowledgeable dialogue system.

Dialogue systems as our chatting partner

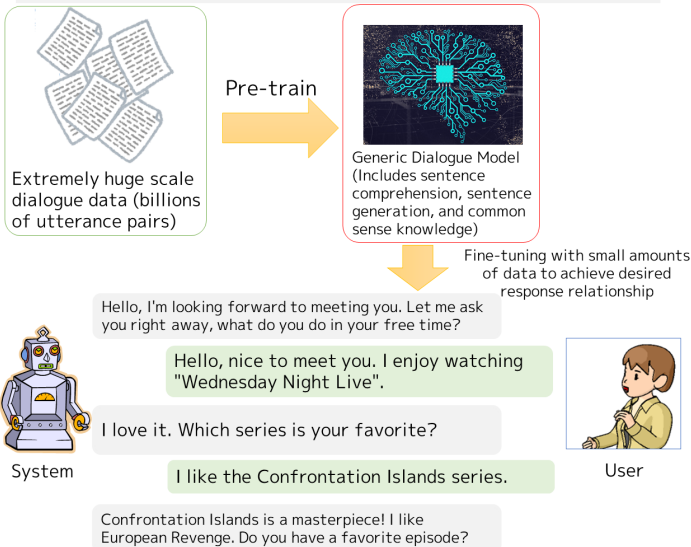The dialogue system is becoming a pleasure partner for chit-chat.
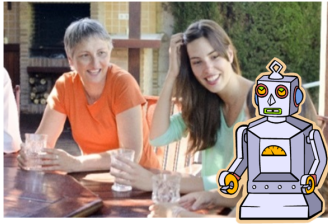
- **Anytime**
  (No restrictions on time or place)
- **Easily**
  (No need to be hesitant)
- **Deeply**
  (Deal with detailed hobbies and discuss private matters)

Dialogue system with pre-training

Large-scale pre-training based methods rapidly improve the performance of chatting systems.

*Pre-training: A method in which the system learns the naturalness of sentences and rough response patterns in advance using large-scale data.

Extremely huge scale dialogue data (billions of utterance pairs)

Pre-train

Generic Dialogue Model (Includes sentence comprehension, sentence generation, and common sense knowledge)

Fine-tuning with small amounts of data to achieve desired response relationship

Hello, I'm looking forward to meeting you. Let me ask you right away, what do you do in your free time?

Hello, nice to meet you. I enjoy watching "Wednesday Night Live".

I love it. Which series is your favorite?

I like the Confrontation Islands series.

Confrontation Islands is a masterpiece! I like European Revenge. Do you have a favorite episode?

System

User

We achieve natural dialogues in text-closed world.
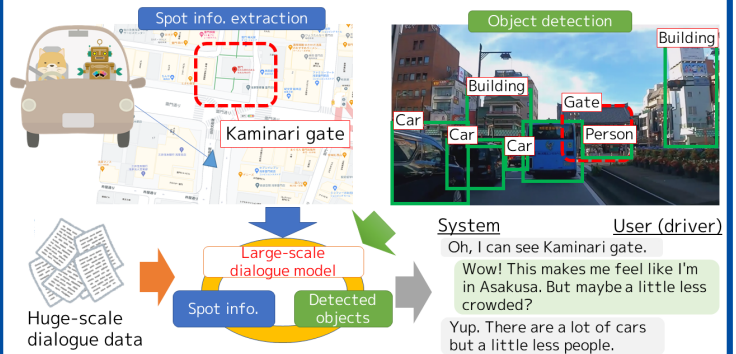
Robot that chats while "watching" the scenery

Problems with text-closed chats
- System cannot interact based on its own surrounding context.
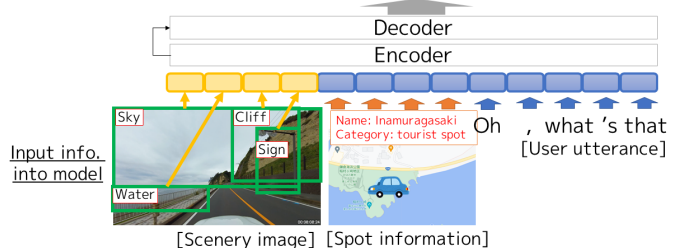
This exhibition
- Realization of a chatting robot that recognizes the ever-changing scenery and geographic information seen from a car while driving.

Spot info. extraction

Object detection

Building

Building

Car

Car

Gate

Car

Person

Kaminari gate

Large-scale dialogue model

Spot info.

Detected objects

Huge-scale dialogue data

System

User (driver)

Oh, I can see Kaminari gate.

Wow! This makes me feel like I'm in Asakusa. But maybe a little less crowded?

Yup. There are a lot of cars but a little less people.

Model Structure
- Introduce image information and external knowledge into text-based chatting system.
- Input image features of detected objects.
- Input textual information on spots around the vehicle's location.
- Selects the most topical utterance from a group of utterances expressing impressions of a sequence of images obtained from video images.

System utterance candidates

1: I heard that area is a tourist spot called Inamuragasaki.
2: That looks like a cliff that could collapse at any moment.

Decoder

Encoder

Sky

Cliff

Sign

Water

Name: Inamuragasaki
Category: tourist spot

Oh, what's that
[User utterance]

Input info. into model

[Scenery image] [Spot information]

References

[1] H. Sugiyama, M. Mizukami, T. Arimoto, H. Narimatsu, Y. Chiba, H. Nakajima, T. Meguro, "Empirical analysis of training strategies of Transformer-based Japanese chit-chat systems," arxiv:2109.05217, 2021.
[2] K. Mitsuda, R. Higashinaka, H. Sugiyama, M. Mizukami, T. Kinebuchi, R. Nakamura, N. Adachi, H. Kawabata, "Fine-tuning a pre-trained Transformer-based encoder-decoder model with user-generated question-answer pairs to realize character-like chatbots," IWSDS, 2021.

Contact

Hiroaki Sugiyama / Interaction Research Group, Innovative Communication Laboratory
Email: cs-openhouse-ml@hco.ntt.co.jp