講演番号 | 164

文献番号 | 20225164

Driving dialogue dataset for a Transformer-based conversational system that talks about scenery

Ko Koga[1)] Toshifumi Nishijima[1)] Hiroaki Sugiyama[2)]

TOYOTA MOTOR CORPORATION, InfoTech Connected Advanced Development Div.,1 Toyota-Cho, Toyota City, Aichi Prefecture 471-8571, Japan (E-mail:ko_koga@mail.toyota.co.jp, E-mail:toshifumi_nishijima@mail.toyota.co.jp)
NTT Communication Science Laboratories, Communication Environment Research Group., KEIHANNA Area and NTT Keihanna Building 2-4, Hikaridai, Seika-cho, "Keihanna Science City" Kyoto, Japan 619-0237 (E-mail:sugiyama.hiroaki@lab.ntt.co.jp)

KEY WORDS: Human, Engineering, Electronics and control, Intelligent, Human Machine Interface [E1]

We would like to develop naturalistic spoken dialog system that can talk about scenery scene, which both brings more enjoyment to travel and prevents distracted driving by drivers. In this paper, we aimed to create a large scenery scene dialogue dataset for fine-tuning such a dialogue system.

First, movies of the driving scenery scene from the driver's point of view was created using the high resolution and high angle of view VR camera (Insta360 Pro 2). Second, we collected videos of several hundred driver-passenger pairs conversing while watching the movies created. We provided the passenger with location information for the route to be traveled, and trained in conversation beforehand. We assumed that the passengers were ideal spoken dialogue system, and passengers were selected from among those who had experience in scriptwriting and customer service from the participants. The number of drivers and passengers was randomly combined, and 750 scene dialogue videos were obtained, in which the drivers and passengers talked about the scenery scene along the Shonan, Yokohama, and Tokyo routes. Third, transcription data was obtained from the scene dialogue videos. Fourth, we completed the development of the Driving Dialogue Dataset, which is a set of scene movies and dialogues, by matching the time of the driving movies and transcriptions (Table 1).

Table 1 Example of Driving Dialogue Dataset

person	start(sec)	end(sec)	utterance	lat	lon
passenger	10.31	11.06	Ann…	35.682722	139.773784
driver	11.42	13.19	Here is Nihonbashi, is'nt it ?	35.683139	139.774034
passenger	17.72	19.34	Hmmm…Nihonbashi?	35.683399	139.774181
driver	19.775	20.555	It's all buildings!	35.683404	139.774184

To characterize the Driving Dialogue Dataset, the following two types of annotations were performed manually by annotators. We defined and categorized three classes related to the scene dialogue, whether the dialogue is talking about the scene you are looking at directly or not. Similarly, we defined and categorized four classes related to self-disclosure of preferences, whether the dialogue is talking about their preferences or not.

The results of the analysis showed that the directly visible scenery scene dialogue was significantly short. Thereafter, the dialogue transitioned to scene related dialogue and scene unrelated dialogue, with longer durations of dialogue. It was found that self-disclosure dialogue of preferences occurs most frequently in driver's speech in the scene unrelated dialogue (Fig.1).
It was also found that the number of self-disclosed utterances of preference varied depending on the driving route such as Shonan, Yokohama, Tokyo.

The scenery scene dialogue suggests that even if the dialogue begins with the scenery scene as the starting point, it is very important that the dialogue system can talk about not only closed domain but also open domain.

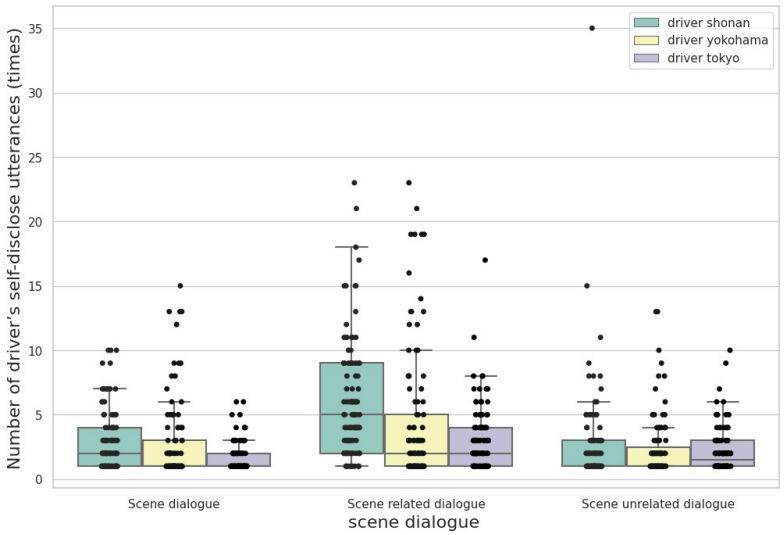

Fig.1 Number of driver's self-disclose utterances of preference for the scene dialogue. (This figure shows that number of driver's self-disclose utterances of preference counted per route and per dialogue session).