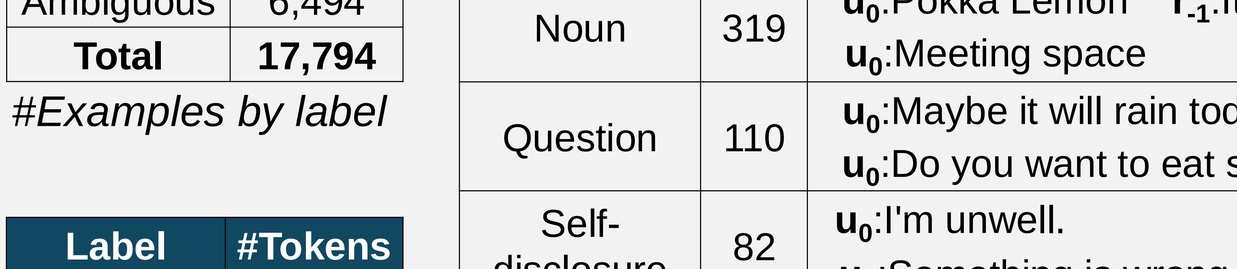
Detecting Ambiguous Utterances in an Intelligent Assistant

Satoshi Akasaki, Manabu Sassano
{sakasaki,msassano}@lycorp.co.jp

Introduction

- Intelligent assistants need to handle a variety of user intents, including tasks and chit-chat
  - ✓ **There are utterances with ambiguous intent**



Related works

- Detect ambiguous utterances in either task-oriented or non-task-oriented systems
  - ✓ Models become specialized for specific settings
- Generating questions that clarify the intent
  - ✓ Which utterances to respond to is unexplored

Dataset construction

- Collected a random sample of **17,794** conversations including user utterances ($\mathbf{u_0}$, $\mathbf{u_{-1}}$, $\mathbf{u_{-2}}$) and system responses ($\mathbf{r_{-1}}$, $\mathbf{r_{-2}}$) as < $\mathbf{u_0}$, $\mathbf{r_{-1}}$, $\mathbf{u_{-1}}$, $\mathbf{r_{-2}}$, $\mathbf{u_{-2}}$ > from the actual intelligent assistant
- Assigned 10 crowd workers per conversation to choose from '**Task**', '**Chat**', or '**Ambiguous**' labels, labeling based on a majority vote of 6 or more.

Label	# Ex.
Chat	5,123
Task	6,177
Ambiguous	6,494
Total	**17,794**

#Examples by label

Label	#Tokens
Chat	4.20
Task	3.85
Ambiguous	2.78

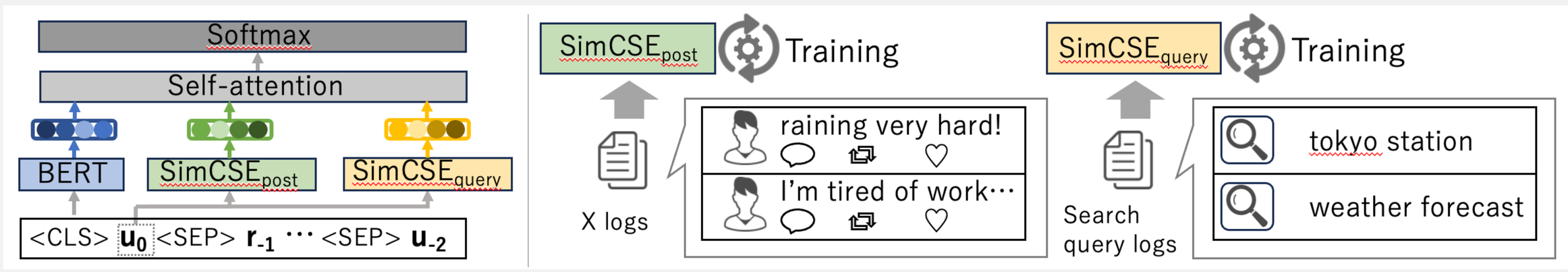
#Avg. tokens by label

Label	# Ex.	Example of conversation (translated from Japanese to English)
Speech Recognition	341	$\mathbf{u_0}$:Today's Tokyo's $\mathbf{r_{-1}}$:Currently, this word is becoming a hot topic... $\mathbf{u_{-1}}$:The latest topic. $\mathbf{u_0}$:Sin. $\mathbf{r_{-1}}$:I sing with all my might. Den-den mushi and katatsumuri... $\mathbf{u_{-1}}$:Sing.
Noun	319	$\mathbf{u_0}$:Pokka Lemon $\mathbf{r_{-1}}$:It seems that the specified app is not installed. $\mathbf{u_{-1}}$:Open Mobage $\mathbf{u_0}$:Meeting space $\mathbf{r_{-1}}$:I, it's troubling to be suddenly told that. $\mathbf{u_{-1}}$:Hug.
Question	110	$\mathbf{u_0}$:Maybe it will rain today. $\mathbf{r_{-1}}$:I hope I have air-conditioning. $\mathbf{u_{-1}}$:It's hot today. $\mathbf{u_0}$:Do you want to eat something? $\mathbf{r_{-1}}$:It's 11:09. $\mathbf{u_{-1}}$:What time is it now?
Self-disclosure	82	$\mathbf{u_0}$:I'm unwell. $\mathbf{r_{-1}}$:You can see the progress here. $\mathbf{u_{-1}}$:Show me. $\mathbf{u_0}$:Something is wrong with my eyes. $\mathbf{r_{-1}}$:You're welcome. $\mathbf{u_{-1}}$:No problem.
Command / Request	104	$\mathbf{u_0}$:I want to go home. $\mathbf{r_{-1}}$:Let's continue the conversation. $\mathbf{u_{-1}}$:What should we talk about? $\mathbf{u_0}$:I want to take a nap. $\mathbf{r_{-1}}$:I am a Voice Assistant. $\mathbf{u_{-1}}$:Tell me your name.
Correction	16	$\mathbf{u_0}$:17:05 $\mathbf{r_{-1}}$:Did you sleep well? The weather in Tokyo is cloudy.. $\mathbf{u_{-1}}$:What time is it now? $\mathbf{u_0}$:The conversation isn't continuing. $\mathbf{r_{-1}}$:Of course. $\mathbf{u_{-1}}$:What are you thinking?
Other	28	$\mathbf{u_0}$:Heeheehee... $\mathbf{r_{-1}}$:I search about ending Puzzle & Dragons. $\mathbf{u_{-1}}$:End `Puzzle & Dragons.' $\mathbf{u_0}$: (・∀・) $\mathbf{r_{-1}}$:With Yahoo! Loco, you can search for various shops and... $\mathbf{u_{-1}}$:Prison School.

Examples of ambiguous utterances

Detection model

- To robustly classify noisy utterances, we feed **sentence embeddings** learned from sources outside the dataset (X posts and queries of a search engine) into the classification model



- Utilize **self-attention** to classify while considering the relationship between embeddings

Experiments and Results

- Evaluate the classification accuracy using the dataset
  - ✓ Models trained with the dataset perform well
- Utterances that belong to 'noun' are difficult to classify
  - ✓ Among nouns, some have ambiguous intents while others do not

Ex:

Shinagawa station ? → Train search / Map search / Web search **VS** **Lexus (car name)** → Web search

	Chat	Task	Ambig.
Threshold [*1]	75.57	80.09	48.62
GPT-4o [*2]	69.72	79.80	55.11
BERT [*3]	80.33	83.73	68.39
Proposed	**82.26**	**84.19**	**71.53**

F_1-score for each label

[*1] Making decisions based on a threshold for binary (chat/task) classification
[*2] Few-shot classification with a prompt
[*3] Fine-tuning with the constructed dataset

Dataset available: https://research.lycorp.co.jp/en/softwaredata