Task proposal for Dialogue System Technology Challenge (DSTC) 6
Dialogue State Tracking on Information Navigation and Attentive Listening

Koichiro Yoshino, Hiroki Tanaka, Sakriani Sakti and Satoshi Nakamura

Graduate School of Informatics, Nara Institute of Science and Technology

8916-5, Takayama-cho, Ikoma, Nara, 6300192, Japan

koichiro@is.naist.jp

1. Background

The task of dialogue state tracking (DST) has been researched as one of the most important components of spoken dialogue systems. In previous dialogue state tracking challenges, DST on typical tasks for spoken dialogue systems, such as the bus navigation system or the restaurant navigation system, were defined as the core shared tasks. Previous dialogue state tracking challenges provided well-organized dialogue frames by hand as the natural language understanding result. However, handcrafting dialogue frames for each task-domain will cost us to build the new tag set and annotations for every new task-domain. The aim of this project is to design a more generalized representation of natural language understanding by limiting the role of dialogue participants. As the first step, dialogue act tagging for information navigation and attentive listening dialogue are proposed. We follow and extend the dialogue act tag set defined in ISO 24617-2 (Bunt et al., 2010). This is the first step to build a general purpose dialogue state tracker, and we will continue to enrich annotations of dialogue data. The task will continue on future DSTCs with the enriched data.

2. Task Description

We propose a task of general purpose dialogue act tagging to capture a dialogue state in the dialogue of information navigation and attentive listening. The dialogue act tag set used in this shared task is an extension of ISO 24617-2 standard dialogue act annotation. The original ISO 24617-2 defines 55 dialogue acts not only for task-oriented dialogue but also for general dialogues, including non-task-oriented dialogue. The original dialogue act tags are classified into 10 classes according to their functions, and acts of information navigation and attentive listening will be merged in this 10 classes:

- Information seeking functions
- Information providing functions
- Commissive functions
- Directive functions
- Auto/Allo feedback functions

- Turn Management functions
- Time Management functions
- Own/Partner Communication Management function
- Discourse Structure Management functions
- Social Obligations Management functions

We will merge the tag sets of information navigation (Yoshino and Kawahara, 2015) and attentive listening (Meguro et al., 2010) to this ISO 24617-2 tag set for this challenge. The information seeking and information providing functions are extended with the tag set of information navigation system (e.g. "inform" act is divided into three classes: "topic-presentation", "story-telling" and "proactive-presentation"). Auto/Allo feedback functions are extended with the tag set of attentive listening system (e.g. "auto-positive" act is divided into "agree" and "empathy"). Some tags defined in these works are merged to the actual classes.

The task of this challenge is automatic tagging of the extended dialogue act tags in transcribed dialogue between two humans. Every dialogue is consist of two phases: information navigation and attentive listening. They talk about some news topics provided by one participant. In the first phase, information navigation part, one participant (navigator) provides a news topic to be discussed. The other attendee (questioner) can ask about the content that the navigator provides, to know the detail of contents according to his or her interest. After the information navigation, the dialogue goes to the second phase, attentive listening. The navigator and the questioner in information navigation change their roles to a listener and a speaker. The speaker declares their opinion about the provided news topic, and the listener listens and sympathizes the opinion of the speaker. Dialogue participants repeat the above procedure for several news topics. Prepared news topics include news that evokes positive and negative emotions.

3. Providing Dataset

3.1. Data

We collected 60 dialogues between an elderly person and listener in Japanese. One dialogue includes 400-500 utterances between participants. We prepared three types of listeners: students, licensed counselor and professional care taker. There are 20 dialogues for each type, and one listener talks with 5 elderly persons. Both of transcription and speech recognition result are provided for each dialogue, and we expect participants to run the system on both inputs. The transcription and the ASR result of Japanese are tokenized with KyTea (Neubig et al., 2011). Translations to English are also provided as the input. Data in the test set are manually translated at least, and some data of training set also will be manually translated, rest of the data are translated in machine translation.

3.2. Providing data example

Information Navigation part

Speaker-ID	DA	Tokenized Utterance (Words, Pronunciations and Translations)
LI-000001	I-G	それ で は 今日 は よろしく お 願 い し ま す 。 s o r e d e w a ky o: w a y o r o sh I k u o n e g a I sh I m a s u <s> Well then, I would like to ask for your help today.
SP-000001	F	(F え) e e
LI-000002	TP-In	(F えっと) 今日 は (F その) (P) 二 つ の ニュース に つ い て 、 ちょっと お 話 し を 二 つ の ニュース に つ い て (F えっと) お 話 し を 伺 っ て い き た い と 思 う ん で す けれ ど も (P) まず 、 この (F ま) 明 る い ニュース で す ね (F えっと) ノーベル 賞 の (F えっと) 授 賞 式 が 十 二 月 頭 に 行 わ れ た の は ご 存 じ で す か ？ (F e t t o) ky o: w a (F s o n o) (P) f u t a t s u n o ny u: s u n i t s u i t e <s> ch o t t o h a n a sh i o f u t a t s u n o ny u: s u i t e (F e t t o) o h a n a sh I o u k a g a q u t e i k I t a i t o o m o u N d e s u k e r e d o m o (P) m a z u <s> k o n o (F m a) a k a r u I ny u: s u d e s u n e (F e t t o) n o: b e r u sy o: n o (F e t t o) jy u sh o: sh i k i g a jy u: n i g a t s u a t a m a n i o k o n a w a r e t a n o w a g o z o n j i d e s u k a <sp> I think that I would like to talk about two news stories about two news today, but first of all, this bright news Do you know that the Nobel Prize awards ceremony was held in the beginning of December?
SP-000002	Ag-AP	(R はい) (R h a i) Yes
SP-000003	Co-In	すん ま せ ん 、 もう 一 度 。 S u N m a s e N <sp> m o: i ch i d o <sp> Sorry?
…		

Attentive Listening part

Speaker-ID	DA	Tokenized Utterance (Words, Pronunciations and Translations)
SP-000051	SO-In	もう 大学 院 お 父 さん 行 か し て 言 う から (F まー) 行 か し て 。
		m o: d a i g a k u i N o t o: s a N i k a sh i t e i u k a r a (F m a) i k a sh i t e
		He said please make me to go to graduate school and I agreed.
LI-000067	Em-AP	(R ふーん) (R うん) (F あっ) (P) そう だ 四 年 前 、 そう で す ね 、 ちょうど もう (P) 就職 氷河 期 真っ 只中 。
		(R f u: N) (R u N) (F a q) (P) s o: d a y o n e N m a e <sp> s o: d e s u n e <sp> ch o: d o m o u (P) sy u: sy o k u hy o: g a k i
		Oh, 4 years before, yes, it was ice ages of job hunting
SP-000052	Ag-In	そう 、 そんな ん な か っ た 。 もう な か っ た 。 四 年 生 で も な か っ た 言 う て 。
		s o u <sp> s o N n a N n a k a t t a <sp> m o u n a k a t t a <sp> y o n e N s e I d e m o n a k a t t a y u: t e
		Yes, he didn't have, didn't have, he said he didn't have even if he was fourth grade undergraduate.
LI-000068	Ag-AP	(R はい)
		(R h a i)
		yes,

...

4. References

I. BUNT, Harry, et al. ISO 24617-2: A semantically-based standard for dialogue annotation. In Proc. The International Conference on Language Resources and Evaluation, 2012. p. 430-437.

II. YOSHINO, Koichiro; KAWAHARA, Tatsuya. Conversational system for information navigation based on POMDP with user focus tracking. Computer Speech & Language, 2015, 34.1: 275-291.

III. MEGURO, Toyomi, et al. Learning to control listening-oriented dialogue using partially observable Markov decision processes. ACM Transactions on Speech and Language Processing (TSLP), 2013, 10.4: 15.

IV. NEUBIG, Graham; NAKATA, Yosuke; MORI, Shinsuke. Pointwise prediction for robust, adaptable Japanese morphological analysis. In: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2. Association for Computational Linguistics, 2011. p. 529-533.