Dialog State Tracking, a Machine Reading Challenge

Julien Perez[1], Ekaterina Kynev[1], and Nicolas Monet[1]

[1]Xerox Research Centre Europe, Grenoble, France

Abstract

We propose a shared task that aims at being an original definition of the setting of dialog state tracking. As the problem is usually framed as a sequential filtering task, we propose to define this task as a machine reading challenge. Among others, we aim at addressing several current challenges faced by dialog state trackers (1) The need for reasoning capable models and associated learning algorithms (2) The need of a human-machine dialogs to be enriched with adversarial inference that off-the-shelf deep learning approaches hardly solve and (3) Variable length dialogs.

1 Overview

The shared task follows the popular trend of machine reading and its recent application to dialog state tracking [Per16]. In this sense, the task will be organized as a set question-answering dataset composed with passage extracted from a corpus of human-machine dialog in the domain of corporate reception. The dialogs have been gathered in the context of an internal research project of human-robot transactional interactions. In this challenge, participants will evaluate their models on a series of dialog state question answering tasks with a different level and nature of complexity. The underlying state is composed with 5 variables over 3 intents which are (1) *Meeting Information Request*, (2) *Parcel Request* and (3) *Delivery*.

2 Relevance

Machine reading has recently received a lot of attention from the Natural Language Processing and Machine Learning communities [HKG+15, WBC+16]. The task of machine reading aims at learning to answer questions about a corpus from triplified examples of {Text, Question, Answer}. To address this challenge a large panel of propositions based on recurrent networks [CDL], attention-based models [HKG+15, PL16] and hybrid systems incorporating featured linguistic representations [CBM16] have been investigated. The challenge aims at bringing all these recent advances to the crucial domain of dialog state tracking. Indeed, as end-to-end dialog systems are desirable due to the relaxation of the data annotation requirement, current systems still rely on the capability of inferring hidden variables and structures from dialog transcriptions. A second aspect of the task concerns the reasoning capabilities we want tracking models to be able to express. Indeed, beyond robust string matching,

we want the model to answer questions about reacher statement that are observed in natural language.

3 Proposed Task

In this challenge, a set of triplified examples of {Text, Question, Answer} will be provided to the participants. Each text corresponds to a passage of a dialog of variable length extracted from a transaction between a user and a robotic receptionist agent. The user utterances will be the 1st-best transcription extracted from a state of the art Automatic Speech Recognition (ASR) module. The question will be produced using semi-automated question generation from the ground-truth state of the dialog. The questions will be organized into four types

Factoïd : this question type takes as support the value of a given variable to track. Figure 1 is an example of such triple.

Binary Factoïd : This question type, also known as *Yes-No* question, asks the assignment of a given variable. An example of such triple is given in Figure 2.

Temporal question : This question type focuses on the capability of temporal reasoning of the proposed model. In the supporting dialog corpus, one variable type corresponds to time in the context of a meeting event. The type of question will correspond to time manipulation. Figure gives an example of such triple.

Binary Composition : In this last question type, binary questions are asked on the basis of a selection of variable assignment available on the passage. Once again, understanding of the question and reading of the dialog will be necessary. Figure 4 is an example of such triple.

4 Setup

We invite submissions of existing and new systems capable of machine reading applied to such dialog tracking problems. A convex metric learning model will be provided as baseline system. A first format of the dataset is described in Appendix A but can be subject to modification and improvement. The dataset will be decomposed into the four types of proposed questions and evaluations will be handled independently from one another. Such setting will allow to analyze the capabilities of the different candidate models on each tasks. The participant are authorized to develop a unique or dedicated model for each question types.

5 Metrics

Despite the complexity of the task, the challenge have a clear evaluation measure. Following existing machine reading tasks, we assume the answer might be either Yes/No or a span of text to extract and lowercase from the dialog passage. An interesting point will be to study how a given model can perform this variety of different tasks, however it will be possible to develop different models for each question type.

References

[CBM16] Danqi Chen, Jason Bolton, and Christopher D. Manning. A thorough examination of the CNN/daily mail reading comprehension task. The Association for Computer Linguistics, 2016.

[CDL] Jianpeng Cheng, Li Dong, and Mirella Lapata. Long short-term

memory-networks for machine reading. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016*, pages 551–561.

[HKG+15] Karl Moritz Hermann, Tomas Kocisky, Edward Grefenstette, Lasse Espeholt, Will Kay, Mustafa Suleyman, and Phil Blunsom. Teaching machines to read and comprehend. In *Proceedings of Advances in Neural Information Processing Systems (NIPS 2015)*, pages 1684–1692, Barcelona, Spain, 2015.

[Per16] Julien Perez. Dialog state tracking, a machine reading approach using a memory-enhanced neural network. *CoRR*, abs/1606.04052, 2016.

[PL16] Julien Perez and Fei Liu. Gated end-to-end memory networks. *CoRR*, abs/1610.04211, 2016.

[WBC+16] Jason Weston, Antoine Bordes, Sumit Chopra, Alexander M Rush, Bart van Merriënboer, Armand Joulin, and Tomas Mikolov. Towards AI-complete question answering: A set of prerequisite toy tasks. In *Proceedings of the 4th International Conference on Learning Representations (ICLR 2016)*, San Juan, Puerto Rico, 2016.

6 Appendix A: Dialog passages and Question Answering Format

```
{
"turn": [
  "Agent: Hello, welcome to Corp headquarter! How can I help you ?",
  "Human: i have a meeting today at eleven thirty my name is john smith",
  "Agent: With whom do you have the appointment ?",
  "Human: johnson"
],
"question": "What is the time of the meeting ?",
"answer": "eleven thirty"
}
```

Figure 1: Example of Factoïd Triple

```
{
"turn": [
  "Agent: Hello, welcome to Corp headquarter! How can I help you?  ",
  "Human: Hi, my name is peter stone and i have a meeting today at five",
  "Agent: Who do you have the appointment with?",
  "Human: i have a meeting with john smith"
],
"question": "Is John the employee's firstname ?",
"answer": "Yes"
}
```

Figure 2: Example of Binary Factoïd Triple

```
{
"turns": [
  "Agent: Hello, welcome to Corp headquarter! How can I help you today?",
  "Human: i know i have a meeting today with mick johnson at eleven thirty",
  "Agent: What company are you from?",
  "Human: i'm from microsoft"
],
"question": "Is the time of the meeting before noon ?",
"answer": "Yes"
}
```

Figure 3: Example of Temporal question

```
{
"turns": [
  "Agent: Hello, welcome to Corp headquarter! How can I help you?  ",
  "Human: morning",
  "Agent: With whom do you have the appointment?",
  "Human: i have a meeting with mister clark",
  "Agent: What company are you from?",
  "Human: i'm from general electric",
  "Agent: What's your name please?",
  "Human: yes my name is peter johnson",
  "Agent: What time is the appointment please?",
  "Human: it is at eleven thirty",
],
"question": " Is the appointment scheduled for eleven thirty with mister clark, the user
    name is peter johnson from general electric?",
"answer": "Yes"
}
```

Figure 4: Example of Binary Composition Question