Tackling Dialog State Tracking Challenge

Wencan Luo
Department of Computer Science
University of Pittsburgh
PA 15260, USA
wencan@cs.pitt.edu

Abstract

We will propose a model to tackle the Dialog State Tracking Challenge (Williams et al., 2013).

1 Introduction

In dialog systems, ”state tracking” sometimes also called ”belief tracking” refers to accurately estimating the users goal as a dialog progresses. Accurate state tracking is desirable because it provides robustness to errors in speech recognition, and helps reduce ambiguity inherent in language within a temporal process like dialog. Dialog state tracking is an important problem for both traditional uni-modal dialog systems, as well as speech-enabled multi-modal dialog systems on mobile devices, on tablet computers, and in automobiles (Williams et al., 2012).

The “Dialog State Tracking Challenge”(DSTC) provided a good test bank for this task. This challenge 2013 has completed. 9 teams entered a total of 27 entries. Results has been shown at SigDial 2013.

The data is still public available [1].

2 Task Description

DSTC data is taken from several different spoken dialog systems. All of them provide bus schedule information for Pittsburgh, Pennsylvania, USA (Black et al., 2013). Different dialog system might have differen ASR, NLU and dialog control components. In this challenge, only 9 slots are evaluated: route, from.desc, from.neighborhood, from.monument, to.desc, to.neighborhood, to.monument, date, and time. The approximate number of distinck values for slots are shown in Table 1. The number of values for each slot varies a lot.

The dialog tracker logs SLU N-best hypotheses for each user turn with confidence scores. As they claimed, the coverage of N-best hypotheses is good, so the challenge confines consideration of goals to slots and values that have been observed in an SLU output. The task of a dialog state tracker is to generate a set of observed slot and value pairs, with a score between 0 and 1. The sum of all scores should be 1.

For evaluation, there are 11 different metric, 4 test tests under 3 different schedules (Williams et al., 2013) for 9 slots.

3 The Corpus

The data is divided into 4 training sets and 4 test sets. They come from different sources. The basic statistical information for the corpus is shown in Table 2

4 Related Work

4.1 Overall Results

4.2 Methodalogy

5 Timeline

Sep 09 - Sep 22

- survey the related work regarding role recognition
- understanding the data, know how to extract and use the data

[1] http://research.microsoft.com/en-us/events/dstc/. This link was broken a few days ago, but it is fixed up after I request

Dataset	Source	Calls	Time period	Transcribed?	Labeled?
train1a	Group A	1013	September 2009	Yes	Yes
train1b	Group A	14,545	16 Months (2008-2009)	Yes	No
train2	Group A	678	Summer 2010	Yes	Yes
train3	Group B	779	Summer 2010	Yes	Yes
test1	Group A	765	Winter 2011-12	Yes	Yes
test2	Group A	983	Winter 2011-12	Yes	Yes
test3	Group B	1037	Winter 2011-12	Yes	Yes
test4	Group C	451	Summer 2010	Yes	Yes

Table 2: Dataset description

Slot name	number of values
route	100
from.desc	500-10000
to.desc	500-10000
from.neighborhood	20-100
to.neighborhood	20-100
from.monument	50-500
to.monument	50-500
date.day	9
date.absmonth	12
date.absday	31
date.relweek	1
time.hour	12
time.minute	60
time.ampm	2
time.arriveleave	2
time.rel	1

Table 1: Approximate number of distinct values for slots

Sep 23 - Oct 20

- implement the method in (**?**) using the manual transcription, the lexical model will be used as the local model
- do Automatical Speech Recognition (ASR)
- run the local model on ASR results

Oct 21 - Nov 9

- implement ILP global model, using the manual speaker segmentation

Nov 10 - Dec 12

- propose a model without the manual speaker segmentation
- try other global model such as Bayes network, improved social network

Acknowledgments

Do not number the acknowledgment section.

References

Jason D Williams, Antoine Raux, Deepak Ramachandran, and Alan W Black. 2012. *Dialog state tracking challenge handbook*. Technical report, Microsoft Research.

A. Black et al. 2011. *Spoken dialog challenge 2010: Comparison of live and control test results*. In Proceedings of SIGDIAL.

Jason D. Williams, Antoine Raux, Deepak Ramachandran, and Alan Black. 2013. *The Dialog State Tracking Challenge*. In Proceedings 14th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL), Metz, France.