Incremental Joint Modelling for Dialogue State Tracking

Anh Duong Trinh, Robert J. Ross, John D. Kelleher
School of Computing
Dublin Institute of Technology
Kevin Street, Dublin 8, Ireland
anhduong.trinh@mydit.ie, {robert.ross, john.d.kelleher}@dit.ie

1 Introduction

Dialogue State Tracking (DST) is a crucial part of Dialogue Systems, as it provides a powerful mechanism to track the user and system's contributions to the dialogue so that the system can determine the best next move in dialogue. In task-oriented Dialogue Systems the distribution over the set of dialogue slots with possible values is called the Dialogue State or State Belief.

While there have been great improvements in DST technology in recent years, there remain two big disadvantages of traditional DST approaches: (1) different DST models are developed separately for different dialogue slots, therefore each model can only partially observe the dialogue; (2) Dialogue States are tracked in a turn-by-turn manner, which lacks flexibility for real-time Spoken Dialogue Systems. The second disadvantage has been recently addressed with LecTrack presented by Zilka and Jurcicek (2015). Aiming to improve on this work, we propose an Incremental Joint Model (IJM) as a novel approach to DST tasks.

2 Incremental Joint Modelling

Generally, dialogues can be treated as a sequence of turns or words, therefore in recent times Recurrent Neural Networks (RNN) have been widely chosen for dialogue tasks. With this in mind, we have developed the IJM tracker, which has the structure shown in Figure 1, based on RNNs with Long Short-Term Memory (Hochreiter and Schmidhuber, 1997).

Our IJM tracker consists of two parts: a shared RNN to handle input and memory channels and separate RNNs to output different components of Dialogue States. We represent words using an embedded vector format and feed these vectors as the input to the network. The memory is a combination of inner RNN memory and previous output in dialogue history. The shared RNN takes into account the input of the current time step and network's memory and produces a universal hidden state. Then the separate RNNs use this universal hidden state to output the probability distribution of particular slots, such as *food* and *price_range*.

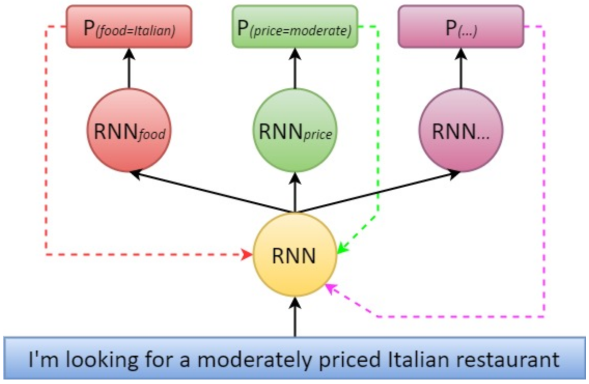

Figure 1: Incremental Joint Modelling tracker. RNN denotes Recurrent Neural Networks, P – Probability Distribution.

The IJM tracker processes dialogues on a word-by-word basis and gives the ultimate output only when it reaches the end of utterance, i.e. when the user stops talking. At one time step only one word is transformed into a vector and put into the network. This incremental manner allows our IJM tracker to produce Dialog States in real time and output them when required.

We have trained and tested the IJM tracker on Dialogue State Tracking Challenge 2 (DSTC2) (Henderson et al., 2014a) data, which has 1612 training, 506 development, and 1117 test dialogues. DSTC2 tasks require trackers to present the Dialogue State consisting of three components for each dialogue turn: Joint Goal Constraints, Search Method and Requested Slots. Trackers' results are evaluated using accuracy metric (Bohus and Rudnicky, 2006) and L2 norm metric (Young

Trackers	Tracker Inputs		Joint Goals		Method		Requested	
	ASR	SLU	Acc.	L2	Acc.	L2	Acc.	L2
Baseline		✓	0.619	0.738	0.879	0.209	0.884	0.196
Web-style ranking & SLU	✓	✓	**0.784**	0.735	0.947	0.087	0.957	0.068
	✓	✓	0.773	0.467	**0.950**	**0.082**	0.968	0.050
Word-based with RNN	✓		0.768	**0.346**	0.940	0.095	**0.978**	**0.035**
LecTrack	✓		0.720	0.640	0.930	0.140	0.970	0.060
Separate Model		✓	0.584	0.779	0.903	0.182	0.954	0.088
Joint Model		✓	0.637	0.658	0.912	0.154	0.954	0.085
Incremental Separate Model	✓		0.702	0.556	0.934	0.124	0.973	0.051
Incremental Joint Model (IJM)	✓		0.707	0.545	0.940	0.114	0.975	0.047

Table 1: Performance of DSTC2 baseline system and best trackers, LecTrack, and our models on DSTC2 test data. Higher accuracy (Acc.) and lower L2 are better.

et al., 2009). Results with higher accuracy and lower L2 norm are better.

3 Results and Discussion

We are currently at an early phase of developing the IJM tracker. However, preliminary evaluation on DSTC2 test data is presented in Table 1. The top four rows of Table 1 present the results of the baseline and best performing systems at the DSTC2 (Henderson et al., 2014a; Williams, 2014; Henderson et al., 2014b), and the state-of-the-art incremental DST LecTrack (Zilka and Jurcicek, 2015), the bottom 4 rows present the results of 4 variants of models we have developed.

Overall, Joint Modelling outperforms Separate Modelling in all tasks, producing higher accuracy and lower L2 norms. Changing input from Spoken Language Understanding (SLU) unit to Auto Speech Recognition (ASR) data, i.e. changing from a turn-by-turn to a word-by-word approach, increases the results substantially. We also found that Joint Modelling trackers outperformed Baseline system provided by the DSTC2 organizers.

The IJM tracker is not competitive yet with best trackers presented in DSTC2, especially in Joint Goals task, which leaves a lot of room to develop our model. Nevertheless, in comparison with the incremental tracker LecTrack, the IJM tracker produces lower accuracy but lower L2 in the Joint Goals task and better results in the Search Method and Requested Slots tasks than LecTrack.

We plan to increase Joint Goals accuracy of our Incremental Joint Model by working on utterance and word vector representations.

Acknowledgments

The authors wish to thank Science Foundation Ireland for their support through the ADAPT Global Centre of excellence for digital content and media innovation.

References

Dan Bohus and Alex Rudnicky. 2006. A K Hypotheses + Other Belief Updating Model. In *Procs. of AAAI Workshop on Statistical and Empirical Methods in Spoken Dialogue Systems 2006*.

Matthew Henderson, Blaise Thomson, and Jason D. Williams. 2014a. The Second Dialog State Tracking Challenge. In *Procs. of the SIGDIAL 2014 Conference*, pages 263–272.

Matthew Henderson, Blaise Thomson, and Steve Young. 2014b. Word-Based Dialog State Tracking with Recurrent Neural Networks. In *Procs. of the SIGDIAL 2014 Conference*, pages 292–299.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation*, 9(8):1735–1780.

Jason D. Williams. 2014. Web-style ranking and SLU combination for dialog state tracking. In *Procs. of the SIGDIAL 2014 Conference*, pages 282–291.

Steve Young, Milica Gasic, Simon Keizer, Francois Mairesse, Jost Schatzmann, Blaise Thomson, and Kai Yu. 2009. The Hidden Information State model: A practical framework for POMDP-based spoken dialogue management. *Computer Speech & Language*, 24(2):150–174.

Lukas Zilka and Filip Jurcicek. 2015. Incremental LSTM-based dialog state tracker. In *Procs. of IEEE Workshop on Automatic Speech Recognition and Understanding, ASRU 2015*, pages 757–762.