RNNS FOR DIALOGUE STATE TRACKING

Ondřej Plátek, Petr Bělohlávek, Vojtěch Hudeček & Filip Jurčíček, PhD[a]
vojta.hudecek@gmail.com, oplatek@ufal.mff.cuni.cz

17th September 2016

Institute of Formal and Applied Linguistics
Faculty of Mathematics and Physics
Charles University in Prague

[a]template by LianTze Lim

INTRODUCTION

(food: ?, area: ?, price: ?)

System: *What type of restaurant are you looking for?*

`request(food)`

(food: ?, area: ?, price: ?)

System: *What type of restaurant are you looking for?*

`request(food)`

User: *I am looking for a japanese restaurant in the city center.*

`inform(food:japanese), inform(area:center)`

(food: japanese, area: center, price: ?)

(food: japanese, area: center, price: ?)

System: *What price range do you prefer?*

`request(price)`

(food: japanese, area: center, price: ?)

System: *What price range do you prefer?*

`request(price)`

User: *I want something cheap.*

`inform(price:cheap)`

(food: japanese, area: center, price: cheap)

WHY TRACK A DIALOGUE

- Dialogue agent tracks the progress compactly
    - Easy update
- Expresses user’s goals and restrictions as well as history

WHAT IS THE DIALOGUE STATE

- Set of pairs `(slot, value)`
- State space of possible combinations
  - Each state described e.g. by Dialog Act Items

    (food: japanese, area: center, price: cheap)
  - Handcrafted ontology - relevant slots

DSTC2 DATASET[1]

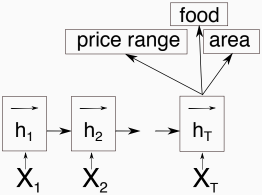
- Challenge organized by University of Cambridge
- Each dialogue composed of *turns*
    - pair of user and system utterances
- Annotated dialogues - true dialogue state after each turn.

- Train set contains 1612, development 506 and test 1117 dialogues
- Dialogues were obtained by user-computer interaction
- Mix of Dialogue Systems and ASR[1] engines was used.
    - Different setting used for each set

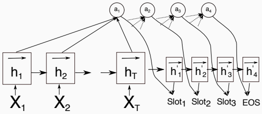
[1] Automatic Speech Recognition

OUR MODEL

- Recurrent neural network composed of sequentially ordered cells (Žilka and Jurčíček [2])
- Capable to process variable length input
- Maintains (encodes) hidden state and emits observation after each turn
- Improved with LSTM cells (Henderson et al. [3])
- Sequence to sequence models adaptation



- Supervised learning
- TensorFlow framework [4] - prepared `RNN, seq2seq`
- Data separated into buckets of similar lengths



FEATURE DETAILS

- INPUT FEATURES
  - feeding input on the world level
  - one-hot encoding (Bag of Words)
  - vector embeddings of words [5]
  - indicator, whether the word belong to particular slot
- We use just 1 best ASR hypothesis
- No SLU[2] employed

[2] Spoken Language Understanding

TRAINING AND EVALUATION

- During training, each turn is predicted based on whole dialogue history
- Metric: **Accuracy** on DS after each turn
  - Schedule 2 - only turns where some information is gained

RESULTS

- Seqeunce-to-sequence model achieved better results - 0.73 vs. 0.727
- Comparable to models that use whole list of ASR hypotheses (Žilka and Jurčíček [2])
- Some systems that use SLU gain better results, up to 0.745 (Henderson et al. [6], Vodolán et al. [7])
- Much better performance of both models on the reshuffled data

- We successfully used the sequence-to-sequence model for DST task
- reasonable performance
- DSTC2 task is much easier on re-shuffled data

FUTURE WORK

- Investigate which features contribute to better performance
- World-level annotated dataset would help to evaluate the incremental models

THE END

THANK YOU!

REFERENCES

[1] M. Henderson, B. Thomson, and J. Williams. The Second Dialog State Tracking Challenge. In *Proceedings of SIGdial*, 2014.

[2] Lukás Žilka and Filip Jurčíček. Incremental LSTM-based dialog state tracker. *arXiv preprint arXiv:1507.03471*, 2015.

[3] M. Henderson, B. Thomson, and S. J. Young. Robust Dialog State Tracking Using Delexicalised Recurrent Neural Networks and Unsupervised Adaptation. In *Proceedings of IEEE Spoken Language Technology*, 2014.

[4] Tensorflow. `https://www.tensorflow.org/`. Accessed: 2016-09-10.

[5] Yoshua Bengio, Rejean Ducharme, and Pascal Vincent. A Neural probabilistic language model. *Journal of Machine Learning Research*, 3:1137–1155, 2003.

[6] Matthew Henderson, Blaise Thomson, and Steve Young. Word-based dialog state tracking with recurrent neural networks. In *Proceedings of the 15th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL)*, pages 292–299, 2014.

[7] Miroslav Vodolán, Rudolf Kadlec, and Jan Kleindienst. Hybrid Dialog State Tracker. *CoRR*, abs/1510.03710, 2015. URL `http://arxiv.org/abs/1510.03710`.