

# Semi-Supervised Bootstrapping of Dialogue State Trackers for Task-Oriented Modelling

Bo-Hsiang Tseng<sup>1,3,5</sup>, Marek Rei<sup>2,4</sup>, Paweł Budzianowski<sup>1,3,5</sup>

Richard E. Turner<sup>1</sup>, Bill Byrne<sup>1</sup>, Anna Korhonen<sup>3</sup>

<sup>1</sup>Engineering Department, Cambridge University, UK

<sup>2</sup>Computer Laboratory, Cambridge University, UK

<sup>3</sup>Language Technology Lab, Cambridge University, UK

<sup>4</sup>Department of Computing, Imperial College London

<sup>5</sup>PolyAI Limited, London, UK

bht26@cam.ac.uk

## Abstract

Dialogue systems benefit greatly from optimizing on detailed annotations, such as transcribed utterances, internal dialogue state representations and dialogue act labels. However, collecting these annotations is expensive and time-consuming, holding back development in the area of dialogue modelling. In this paper, we investigate semi-supervised learning methods that are able to reduce the amount of required intermediate labelling. We find that by leveraging un-annotated data instead, the amount of turn-level annotations of dialogue state can be significantly reduced when building a neural dialogue system. Our analysis on the MultiWOZ corpus, covering a range of domains and topics, finds that annotations can be reduced by up to 30% while maintaining equivalent system performance. We also describe and evaluate the first end-to-end dialogue model created for the MultiWOZ corpus.

## 1 Introduction

Task-oriented dialogue models aim at assisting with well-defined and structured problems like booking tickets or providing information to visitors in a new city (Raux et al., 2005). Most current industry-oriented systems rely on modular, domain-focused frameworks (Young et al., 2013; Sarikaya et al., 2016), with separate components for user understanding (Henderson et al., 2014), decision making (Gašić et al., 2010) and system answer generation (Wen et al., 2015). Recent progress in sequence-to-sequence (seq2seq) modelling has enabled the development of fully neural end-to-end architectures, allowing for different components to be optimized jointly in order to share information (Wen et al., 2017; Zhao et al., 2017; Budzianowski and Vulić, 2019).

Dialogue systems benefit greatly from optimizing on detailed annotations, such as turn-level dia-

logue state labels (Henderson et al., 2014) or dialogue actions (Rieser and Lemon, 2011), with end-to-end architectures still relying on intermediate labels in order to obtain satisfactory results (Lei et al., 2018). Collecting these labels is often the bottleneck in dataset creation, as the process is expensive and time-consuming, requiring domain and expert knowledge (Asri et al., 2017). Due to this restriction, existing datasets for task-oriented dialogue are several orders of magnitude smaller compared to general open-domain dialogue corpora (Lowe et al., 2015; Henderson et al., 2019).

Arguably, one of the most challenging parts of dialogue modelling is maintaining an interpretable internal memory over crucial domain concepts (Williams et al., 2016). Although there is increasing research effort to learn Dialogue State Tracking (DST) jointly with the text generation component (Eric et al., 2017; Wu et al., 2019), the most effective models use it as an intermediate signal (Wen et al., 2017; Lei et al., 2018). The difficulty of state tracking has made this task a driving force behind most of the Dialog System Technology Challenges in recent years (Henderson et al., 2014; Kim et al., 2017).

In this paper, we reduce the reliance of task-oriented dialogue systems on data collection by leveraging semi-supervised training (Chapelle et al., 2009). Two approaches are investigated and evaluated for providing an improved training signal to the dialogue state tracking component in an end-to-end dialogue system. Automatically predicted DST output on unlabelled utterances is treated as additional annotation if the model confidence is sufficiently high. Furthermore, subtle perturbations of existing datapoints are created, optimizing for their predictions to be similar to the original instances. Our analysis on the MultiWOZ corpus (Budzianowski et al., 2018), covering a range of domains and topics, finds that these meth-

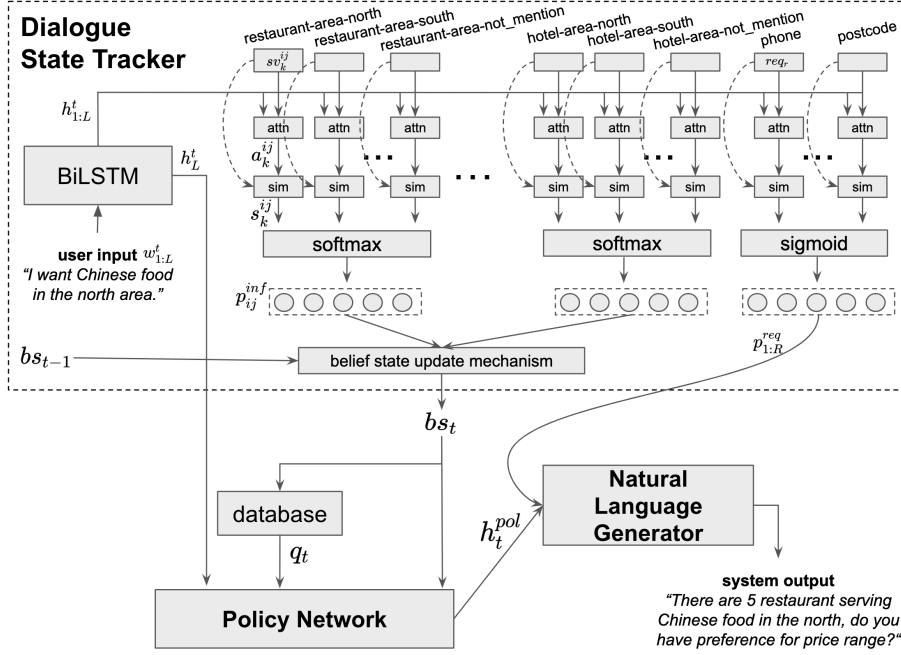


Figure 1: Overview of our end-to-end neural dialogue model. It is composed of three main components: Dialogue State Tracker, Policy Network and Natural Language Generator.

ods can reduce intermediate annotation by up to 30% while maintaining equivalent system performance. We also describe and evaluate the first end-to-end dialogue model created for the MultiWOZ corpus.

## 2 End-to-end Neural Dialogue Model

We now present the end-to-end neural dialogue model composed of three main components: dialogue state tracker, policy network and natural language generator. These components are trained jointly as a connected network and will be introduced in detail in the next paragraphs. The overall architecture can be seen in Figure 1.

**Dialogue State Tracker (DST)** The DST is responsible for both understanding the input user utterance and updating the internal dialogue state for the downstream components. There are two types of slots which can be detected in the input: informable slots and requestable slots. The former describes the attributes of the entity that the user is looking for, e.g., `pricerange` of a hotel. The latter captures information that the user desires to know about the entity, e.g., `postcode` of a hotel. Each informable slot contains several possible values with two special labels: `not-mentioned` and `don't-care`. A good DST should be able to correctly recognize the mentioned slot-value pairs in the user utterance and to maintain the updated

dialogue (belief) state.

Let  $i$ ,  $j$  and  $k$  denote the index of domain, slot and value. As depicted at the top of Figure 1, the user utterance  $w_1:w_L$  at turn  $t$  is first encoded by the BiLSTM to obtain the hidden states  $h_{1:L}^t$ . The encoding of the slot-value pair  $sv_k^{ij}$  is the output of the affine layer that takes the concatenation of the embeddings of domain  $i$ , slot  $j$  and value  $k$  as the input. The context vector  $a_k^{ij}$  is then computed by the attention mechanism, denoted as `attn` in Figure 1, following Luong et al. (2015):

$$e_l = \text{sim}(h_l, sv_k^{ij}) \quad (1)$$

$$a_k^{ij} = \sum_{l=1}^L e_l h_l, \quad (2)$$

where  $l$  is the word index of the user utterance and `sim` denotes any function that calculates the similarity of two vectors. We adopt here the dot product function, following Mrkšić et al. (2017); Zhong et al. (2018); Ramadan et al. (2018). The similarity score  $s_k^{ij}$  between  $a_k^{ij}$  and  $sv_k^{ij}$  is then computed to see whether the slot-value pair  $sv_k^{ij}$  is mentioned in the utterance. The mentioned pair should have higher similarity score to its context vector than those which are not mentioned. The softmax layer is then applied to form the probability distribution  $p_{ij}^{inf}$  for each informable slot  $s_{ij}^{inf}$ , where the predicted value is the value with

the highest probability. The same attention mechanism is used for each requestable slot  $req_r$  to decide whether the user has asked for the slot in the current turn. The sigmoid layer is used instead as it is a binary classification problem. The prediction of requestable slots will be used as input to the natural language generator.

The belief state is the concatenation of the distributions over all informable slot-value pairs that is updated at each turn to keep track of the information provided by the user during the entire conversation. To form the belief state  $bs_t$  at turn  $t$ , for each informable slot  $s_{ij}^{inf}$  the update mechanism checks if the predicted value is either `not-mention` or `dont-care` for the current turn. If it is, then the probabilistic distribution  $p_{ij}^{inf}$  in  $bs_{t-1}$  is kept, otherwise it is updated by the new distribution  $p_{ij}^{inf}$  at the current turn.

**Policy Network** The policy network is responsible for fusing the signals from the belief state  $\mathbf{b}_t$ , the encoding of the user utterance  $\mathbf{h}_L^t$  and the database query result  $\mathbf{q}_t$ . The database query is constructed from the predicted belief state. The number of all entities that match the predictions of the DST form the database query vector<sup>1</sup>. We use a simple feedforward layer as the policy network:

$$\mathbf{z}_t = \tanh(W^z * [\mathbf{b}_t, \mathbf{h}_L^t, \mathbf{q}_t]). \quad (3)$$

where  $[*]$  denotes the concatenation of vectors.

**Natural Language Generator** Taking the input  $\mathbf{z}_t$  from the policy network and predictions of requestable slots from the tracker, the generator outputs a system response word-by-word until the `<EOS>` token is generated. To improve the generation of correct slots corresponding to the user input, we adopt the semantically-conditioned LSTM (Wen et al., 2015) that contains a self-updating gate memory to record the produced slots in the generated sentence.

**Optimization** The model is optimized jointly against two sources of information – DST intermediate labels and system utterances. The DST loss consists of the cross-entropy over the multi-class classification of informable slots while bi-

<sup>1</sup>Following (Wen et al., 2017), by querying the database using  $\mathbf{b}_t$  as query,  $\mathbf{q}_t$  is the 1-hot representation with each element indicating different number of the matching entities in the database. We use 5 bins to indicate the matching number from 0 to 3 and more than 3.

nary cross-entropy is used for requestable slots:

$$L_{dst} = - \sum_i \sum_j t_{ij}^{inf} \log p_{ij}^{inf} - \sum_r t_r^{req} \log p_r^{req}, \quad (4)$$

where  $i$ ,  $j$  and  $r$  are the index of domain, informable slot and requestable slot respectively;  $t^*$  is the target distribution. The generation error is a standard cross-entropy between the predicted words and target words:

$$L_{gen} = - \sum_{l'} t_{l'} \log p_{l'}, \quad (5)$$

where  $l'$  is the word index in the generated sentence. To jointly train the DST, the policy network, and the generator as a connected network, the final objective function becomes  $L = L_{dst} + L_{gen}$ .

**Semi-supervised Training** The DST loss requires each turn to be manually annotated with the correct slot-value pairs. We experiment with two different semi-supervised training methods that take advantage of unlabelled examples instead, allowing us to reduce the amount of required annotation. The first approach is based on the *pseudo-labelling* strategy by Chapelle et al. (2009). If the prediction probability of an unlabelled data point for a particular class is larger than a given threshold  $\nu$ , the example is included in the DST loss with the predicted label. The  $\nu$  parameter is optimized on the validation set during development.

The second semi-supervised technique investigated is the *II-model* (Sajjadi et al., 2016) where the input is perturbed with random noise  $\epsilon \sim \mathcal{N}(\mathbf{0}, \sigma)$ . The perturbations are applied to both labelled and unlabelled data points at the level of embedding of user utterance. The model is then required to produce similar predictions  $p_{ij}^{inf}$  over the belief state compared to the original input, optimized with an additional loss:

$$L3 = \alpha \frac{1}{N} \sum_N \sum_{ij} (t_{ij}^{inf} - p_{ij}^{inf})^2, \quad (6)$$

where  $N$  is the batch size and  $\alpha$  is a hyperparameter controlling the weight of the loss.

### 3 Experiments

We investigate the effects of semi-supervised training on optimizing the end-to-end neural dialogue system. In particular, we evaluate how much annotation at the intermediate-level could be reduced while preserving comparable results of the overall dialogue task completion metrics.

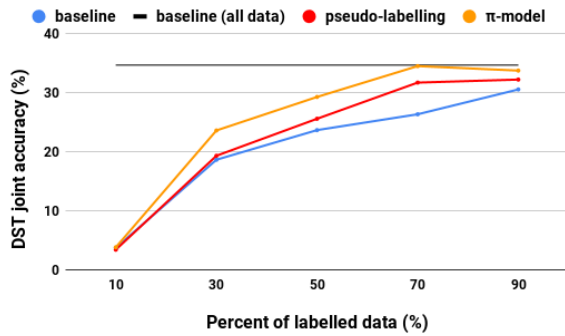


Figure 2: The DST joint accuracy for the three considered models as the amount of labelled data varies. The horizontal line denotes the baseline model trained on 100% labelled data.

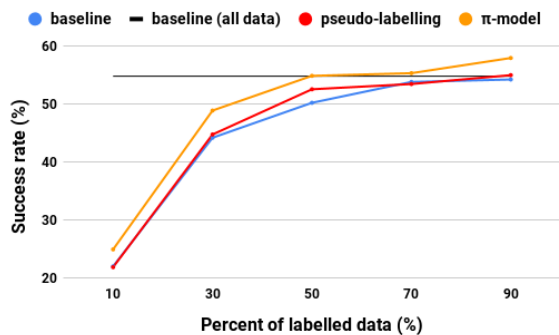


Figure 3: Success rate for different methods as the amounts of labelled data varies. Horizontal line denotes the baseline model trained on 100% labelled data.

**Dataset** The three analyzed models are evaluated on the MultiWOZ dataset consisting of 10,438 task-oriented dialogues (Budzianowski et al., 2018). The conversations in MultiWOZ are natural as they were gathered based on human-to-human interactions following the Wizard-of-Oz paradigm. The corpus includes multi-domain conversations spanning across 7 domains including Restaurant, Hotel, Attraction, Train and Taxi. The size of the dataset allows us to control the amount of available fully labelled datapoints.

**Metrics** There are two metrics of importance when evaluating task-oriented dialogue systems. The first is the DST joint goal accuracy, defined as an average joint accuracy over all slots per turn (Williams et al., 2016). The second is the *Success* metric that informs how many times systems have presented the entity satisfying the user’s goal and provided them with all the additional requested information (Wen et al., 2017). The models are optimized using the validation set and the results are averaged over 10 different seeds.

**Varying data amount** We examine the performance of the baseline model compared to the two semi-supervised models as the amount of labelled data varies. The result of the DST joint accuracy is presented in Figure 2. The *pseudo-labelling* model performs better than the baseline when more than 50% of the dataset is labelled. At the scarce data levels (10% and 30%), the *pseudo-labelling* model is not producing pseudo training points that help improve DST predictions. In contrast, the  $\Pi$ -model takes advantages of the additional regularization loss and effectively leverages unlabelled data to enhance the performance over the baseline. The improvements are consistently more than 5% when training with 30 to 90% of labelled data and even reach the performance of the fully trained baseline model with only 70% labelled data.

Figure 3 shows the *Success* metric results. The *pseudo-labelling* method is not able to improve performance over the baseline regardless of the amount of labelled data. However, the  $\Pi$ -model is capable of improving the success rate consistently and manages to reach the performance of the fully trained model with only 50% of the intermediate DST signal. Note that a better DST joint accuracy does not necessarily translate to a better success rate as the final metric is also influenced by the quality of the generator.

No. of examples	0	1-5	6-10	10-15	16-20
<b>Baseline</b>	6.17	15.93	25.71	35.07	28.88
<b>Pseudo-labelling</b>	6.5	16.27	26.96	33.46	28.55
<b><math>\Pi</math>-model</b>	<b>6.6</b>	<b>21.93</b>	<b>31.29</b>	<b>36.22</b>	<b>30.72</b>

Table 1: The accuracy (%) of different classification of slot-value pairs in terms of their number of training examples.

**DST analysis** DST joint accuracy considers all slot-value pairs in an utterance and cannot give us further insight regarding the source of the improvements. We are particularly interested in whether the semi-supervised models can leverage unlabelled data to improve the prediction of rarely seen slot-value pairs. In this analysis, we classify all slot-value pairs in the test set in terms of their number of training examples in 50% of the labelled data. Table 1 presents the results, showing that the  $\Pi$ -model improves accuracy by 5% when the slot-value pair is rarely (1-10 times) seen during training. The improvement on few-shot slots contributes to the improvement of joint accuracy.



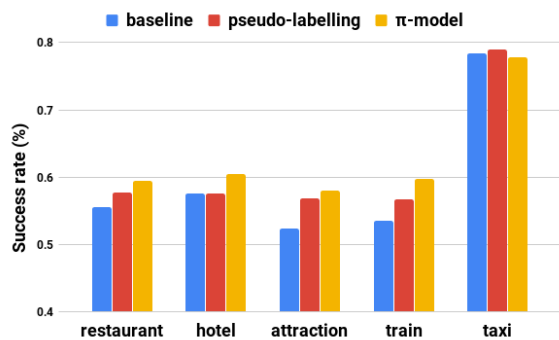


Figure 4: Success rates in each domain in the case of 50% labelled data.

**Domain analysis** We also investigate if the improvements in the success rate are consistent among all domains. Figure 4 shows the success rate on individual domains in the case of 50% of data is labelled. Both semi-supervised models improve performance over the baseline in all domains except for the taxi domain. We hypothesize this comes from the fact that the taxi domain is a relatively easy domain with only 4 possible slots.

## 4 Conclusions and Future Work

In this paper, we have analyzed how much semi-supervised techniques could help to reduce the need for intermediate-level annotations in training neural task-oriented dialogue models. The results suggest that we do not need to annotate all intermediate signals and are able to leverage unannotated examples for training these components instead. In the future, we plan to experiment with other intermediate signals like dialogue acts. Further improvements could potentially be obtained from employing more advanced regularization losses (Oliver et al., 2018).

## Acknowledgments

Bo-Hsiang Tseng is supported by Cambridge Trust and the Ministry of Education, Taiwan. Marek Rei’s research is supported by Cambridge English via the ALTA Institute. Paweł Budzianowski is funded by an EPSRC grant (EP/M018946/1) and by Toshiba Research Europe Ltd, Cambridge Research Laboratory (RG85875).

## References

Layla El Asri, Hannes Schulz, Shikhar Sharma, Jeremie Zumer, Justin Harris, Emery Fine, Rahul Mehrotra, and Kaheer Suleman. 2017. *Frames: A*

corpus for adding memory to goal-oriented dialogue systems. *Proceedings of SigDial*, pages 207–219.

Paweł Budzianowski and Ivan Vulić. 2019. Hello, it’s GPT-2 - how can I help you? towards the use of pre-trained language models for task-oriented dialogue systems. *arXiv preprint arXiv:1907.05774*.

Paweł Budzianowski, Tsung-Hsien Wen, Bo-Hsiang Tseng, Iñigo Casanueva, Stefan Ultes, Osman Ramadan, and Milica Gasic. 2018. *MultiWOZ-A Large-scale multi-domain wizard-of-oz dataset for task-oriented dialogue modelling*. In *Proceedings of EMNLP*, pages 5016–5026.

Olivier Chapelle, Bernhard Scholkopf, and Alexander Zien. 2009. *Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]*. *IEEE Transactions on Neural Networks*, 20(3):542–542.

Mihail Eric, Lakshmi Krishnan, Francois Charette, and Christopher D. Manning. 2017. *Key-value retrieval networks for task-oriented dialogue*. In *Proceedings of SigDial*, pages 37–49.

Milica Gašić, Filip Jurčićek, Simon Keizer, François Mairesse, Blaise Thomson, Kai Yu, and Steve Young. 2010. *Gaussian processes for fast policy optimisation of pomdp-based dialogue managers*. In *Proceedings of SigDial*, pages 201–204.

Matthew Henderson, Paweł Budzianowski, Iñigo Casanueva, Sam Coope, Daniela Gerz, Girish Kumar, Nikola Mrkšić, Georgios Spithourakis, Pei-Hao Su, Ivan Vulić, and Tsung-Hsien Wen. 2019. *A repository of conversational datasets*. In *Proceedings of the Workshop on NLP for Conversational AI*.

Matthew Henderson, Blaise Thomson, and Jason D Williams. 2014. *The third dialog state tracking challenge*. In *Spoken Language Technology Workshop (SLT), 2014 IEEE*, pages 324–329. IEEE.

Seokhwan Kim, Luis Fernando DHaro, Rafael E Banchs, Jason D Williams, and Matthew Henderson. 2017. *The fourth dialog state tracking challenge*. In *Dialogues with Social Robots*, pages 435–449. Springer.

Wenqiang Lei, Xisen Jin, Min-Yen Kan, Zhaochun Ren, Xiangnan He, and Dawei Yin. 2018. *Seqicity: Simplifying task-oriented dialogue systems with single sequence-to-sequence architectures*. In *Proceedings of ACL*, pages 1437–1447.

Ryan Lowe, Nissan Pow, Iulian V Serban, and Joelle Pineau. 2015. *The ubuntu dialogue corpus: A large dataset for research in unstructured multi-turn dialogue systems*. In *16th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, page 285.

Thang Luong, Hieu Pham, and Christopher D Manning. 2015. *Effective approaches to attention-based neural machine translation*. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1412–1421.

- Nikola Mrkšić, Diarmuid Ó Séaghdha, Tsung-Hsien Wen, Blaise Thomson, and Steve Young. 2017. [Neural belief tracker: Data-driven dialogue state tracking](#). In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 1777–1788.
- Avital Oliver, Augustus Odena, Colin A Raffel, Ekin Dogus Cubuk, and Ian Goodfellow. 2018. [Realistic evaluation of deep semi-supervised learning algorithms](#). In *Advances in Neural Information Processing Systems*, pages 3235–3246.
- Osman Ramadan, Paweł Budzianowski, and Milica Gašić. 2018. [Large-scale multi-domain belief tracking with knowledge sharing](#). In *Proceedings of ACL*, pages 432–437.
- Antoine Raux, Brian Langner, Dan Bohus, Alan W Black, and Maxine Eskenazi. 2005. [Let’s go public! taking a spoken dialog system to the real world](#). In *Ninth European Conference on Speech Communication and Technology*.
- Verena Rieser and Oliver Lemon. 2011. *Reinforcement learning for adaptive dialogue systems: a data-driven methodology for dialogue management and natural language generation*. Springer Science & Business Media.
- Mehdi Sajjadi, Mehran Javanmardi, and Tolga Tasdizen. 2016. [Regularization with stochastic transformations and perturbations for deep semi-supervised learning](#). In *Advances in Neural Information Processing Systems*, pages 1163–1171.
- Ruhi Sarikaya, Paul A Crook, Alex Marin, Minwoo Jeong, Jean-Philippe Robichaud, Asli Celikyilmaz, Young-Bum Kim, Alexandre Rochette, Omar Zia Khan, Xiaohu Liu, et al. 2016. [An overview of end-to-end language understanding and dialog management for personal digital assistants](#). In *2016 IEEE Spoken Language Technology Workshop (SLT)*, pages 391–397. IEEE.
- Tsung-Hsien Wen, Milica Gašić, Nikola Mrkšić, Pei-Hao Su, David Vandyke, and Steve Young. 2015. [Semantically conditioned lstm-based natural language generation for spoken dialogue systems](#). In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.
- Tsung-Hsien Wen, David Vandyke, Nikola Mrkšić, Milica Gašić, Lina M Rojas-Barahona, Pei-Hao Su, Stefan Ultes, and Steve Young. 2017. [A network-based end-to-end trainable task-oriented dialogue system](#). In *Proceedings of EACL*, pages 438–449.
- Jason Williams, Antoine Raux, and Matthew Henderson. 2016. [The dialog state tracking challenge series: A review](#). *Dialogue & Discourse*, 7(3):4–33.
- Chien-Sheng Wu, Richard Socher, and Caiming Xiong. 2019. [Global-to-local memory pointer networks for task-oriented dialogue](#). *ICLR*.
- Steve Young, Milica Gašić, Blaise Thomson, and Jason D. Williams. 2013. [POMDP-based statistical spoken dialog systems: A review](#). *Proceedings of the IEEE*, 101(5):1160–1179.
- Tiancheng Zhao, Ran Zhao, and Maxine Eskenazi. 2017. [Learning discourse-level diversity for neural dialog models using conditional variational autoencoders](#). In *Proceedings of ACL*, pages 654–664.
- Victor Zhong, Caiming Xiong, and Richard Socher. 2018. [Global-locally self-attentive encoder for dialogue state tracking](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1458–1467.