International Journal Of Engineering Research And Development
e- ISSN: 2278-067X, p-ISSN: 2278-800X, www.ijerd.com
Volume 15, Issue 3 (March 2019), PP.23-28

Dialogue State Tracking Accuracy Enhancement by Distinguishing Candidate Slot-Value Pairs

Khaldoon H. Alhussayni[1], Alexander Zamyatin[2], Ghassan Khazal[2]
[1]*Computer Center, Babylon University, Babylon, Iraq.*
[2]*ComputerScience Department, Tomsk State University, Tomsk, Lenin Avenue 36, Russia*
Corresponding Author: Khaldoon H. Alhussayni

ABSTRACT: *Dialogue state tracking (DST) playsa critical role ina task-oriented dialogue system's cycle life. DST follows the goals of the user at each turn through dialogue and summarizes these goals as a semantic frame containing slot-value pairs and dialogue acts, which directly affect the performance and effectiveness of dialogue systems. There are different challenges in DST such as linguistics diversity, dynamic context, and distribution of the dialogue state over candidate values in both slot-value and dialogue acts that are defined in the ontology.In this paper,we focus on these challenges by combining current and previous user utteranceto figure the distribution of the slot-value pairs and dialogue acts to increase the performance. The WoZ dataset was used for evaluating the proposed model;the implementationof a two-variant was attempted, first by using previous user utterance as an additional encoder in the dialogue and, second, by using the additional score that combines the context of previous user utterance and current user utterance with all candidate slot-value pairs. The proposed model achieved outperforming results compared with all the state-of-the-art approaches in the joint goal accuracy by 0.8%, but that is not in the request turn task.*

Date of Submission: 25-08-2019 Date of acceptance: 09-09-2019

I. INTRODUCTION

Dialoguestate tracking (DST) is a key component of task-orienteddialogue systems whose responsibility is to keep track of a user's goals at each turn based on the history of the dialogue. Task-oriented dialogue systems interact with users in natural language by users speaking to a voice interface or writing a text to accomplish tasks they have in mind. State-of-the-art approaches for DST depend on deep learning models.

Several neural-based DST systems have recently been proposed. (Mrkšić, Séaghdha, Wen, Thomson, & Young, 2016)proposed the neural belief tracker (NBT) model to apply representationlearning usingthe deep network (DNN) and convolutional network (CNN)to calculate such representation vectors tolearn features appropriate for each state; this is different from handcrafting features, where representation is computed based on pretrained word embeddings to agree with the richness in natural language.However, multivalued slots do not consider all the values, whereas this work can predict probabilities for multiple possible values(Zhong, Xiong, & Socher, 2018).Zhong et al. (2018) proposed the global-locally self-attentive encoder (GLAD)model withself-attention-based recurrent networks for each utterance and previous system actions andcomputed representation by measuring similarity toeach slot-value(Nouri & Hosseini-Asl, 2018). This improved GLAD construction by removing slot-dependent recurrent networks for utterance and a system action encoder and employing a globallyconditioned encoder (GCE)on the slot type embedding vector. However, in the last two approaches, they have still been ineffective for distribution in the production system due to their inactivity with recognizing and incorporating the applicable context, whereas this work can identify the related context.DST aims to predict the set of goals that references ontology items in user utterances, which are represented as slot-value pairs. This becomes a complex mission when challenged with a lexical difference, the dynamics of context output, the distribution of the dialogue state, overall candidate slot-value pairs, and dialogue acts that are defined in the ontology.

In this work, we introduce an enhancement model that was utilized in by(Nouri & Hosseini-Asl, 2018), using the previous slots values that compose the state during each turn. To do this, we combined the current and previous user utterance and measured similarities between them and the ontology.

II. RELATED WORK

Dialogue state tracking methods can be divided into a rule-based approach, statistical approach and deep learning approach. use rule-based heuristics and compute the confidence scores of the N-best candidates generated to inform appropriate dialog states from the output of a natural language understanding module (Higashinaka, Nakano, & Aikawa, 2003; J. Williams, A. Raux, D. Ramachandran, 2017; K. Sun, L. Chen, S.

Zhu, 2014; Traum, 2000) to track the necessary information for tracking a dialog state. However, these rules are not automatically obtained from actual dialogue information so that careful tuning and sensitive design attempts are required. Because of these techniques often lead to inaccuracy in the determination of dialogue states.

As an alternative to handcrafted rules, statistical methods have been utilized to dialogue state tracking (Bohus & Rudnicky, 2006; Ma, Raux, Ramachandran, & Gupta, 2012; Thomson & Young, 2010). Statistical methods such as logistic regression and the Bayesian network for achieving high tracking performance and initiating an assessment of the confidence of the user's information. However, such studies share a prevalent issue that all possible dialogue states must be listed, which is very costly computing.

Recently, using deep learning techniques of dialog state tracking (Abhinav Rastogi , Dilek Hakkani-T ̈ ur, 2017; Casanueva et al., 2017; Henderson, Thomson, & Young, 2013, 2014b, 2014a; Mrkšić et al., 2015, 2016; Perez & Liu, 2016; Wen et al., 2016; Zilka & Jurcicek, 2016) and others can understand particular depictions for user and system utterances and prior system actions to predict the turn state.

The neural network was first used for dialogue state tracking by (Henderson et al., 2013). Their research is important because the first attempt is to use a neural network to dialogue state tracking in a pipeline strategy to collect appropriate data from the user utterances. In the lack of a necessary dialogue framework needed for user interpretation, these schemes are susceptible to error accumulation in the speechlanguage understanding module separately.

The outcomes of the State Tracking Dialog show that it is useful to jointly learn speech understanding and dialogue tracking (Henderson et al., 2014b; Wen et al., 2016; Zilka & Jurcicek, 2016). The N-Best list generated by the automatic speech recognition scheme is the source of these approaches. By preventing error accumulation from the original speech language understanding element. These models rely on delexicalization, using generic tags to substitute particular slot kinds and values, and handcrafted semantic dictionaries. However, these models are depending on handcrafted characteristics and complicated domainspecific lexicons, which are hard to scale for each type of slot and therefore difficult to extend to new domains.

recent state-of-the-art models for DST predict the state of each turn by learning universal representations for user and system utterances. However, the performance of these schemes is poor in rare and unknown slot values, which have lately been handled by local slot encoders (Nouri & Hosseini-Asl, 2018; Zhong et al., 2018) and by the pointer network (Xu & Hu, 2018).

In (Zhong et al., 2018), the global-local self-attention-encoder model has suggested self-attention-recurring networks with a computed representation by measurement of the resemblance of each slot value for each user utterance and prior system actions. (Nouri & Hosseini-Asl, 2018)the enhanced global-local self-attention-encoder structure through the removal of slot-based recurring utterance networks and system encoder, and the use of an embedded slot type global-conditioned encoder.However, these methods were not effective to distribute in the manufacturing scheme, because of their lack of activity in acknowledging and integrating the relevant context, while this work could recognize the associated context.

III. PROPOSED MODEL

In this section, we describe the proposed model. First, section 3.1 describes the recently proposed GCE(Nouri & Hosseini-Asl, 2018)model architecture, followed by proposed encoder in section 3.2, then scour model in section 3.3.

3.1. GloballyConditioned Encoder (GCE)

Each input of encoders is represented as a vector representation(C). A GCE(Nouri & Hosseini-Asl, 2018)model employs bidirectionalLSTM (Sepp & Jurgen, 1997) and all slots that are shared across to obtain a sequence of hidden states (H) by encoding the inputssequence, shadowed by a self-attention layer(Lin et al., 2017)to extract the context vectorand predict the probability distribution of each slot over values. The GCEapproach considersuser utterance and the previous system actiontolearn model. However**,** in our work, we employeda GCEapproach to learning slot-value pairs distribution and context representation,not only current user utterance andprevious system actions but also for previous user utterance.

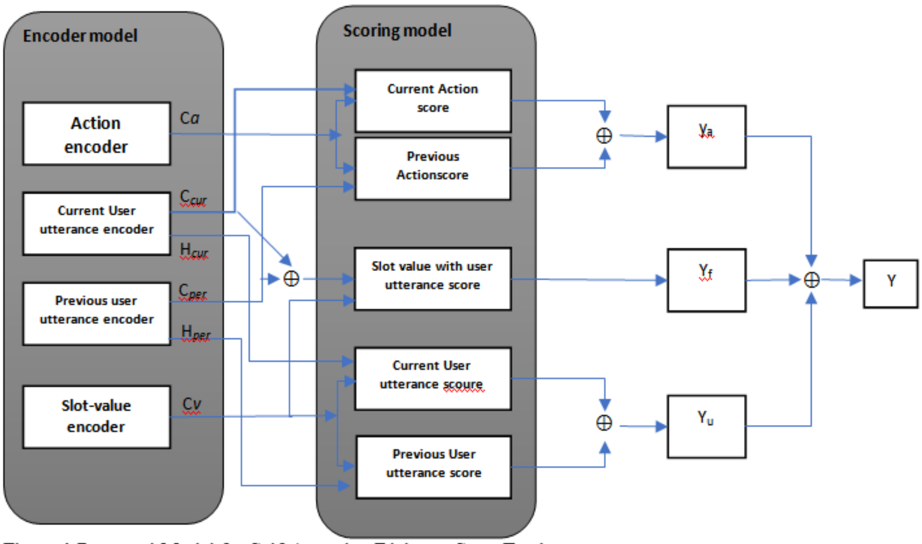

Figure 1.Proposed Model for Self-Attentive Dialogue State Tracker

3.2. Encoder Model

We followed the proposed architecture in GCE(Nouri & Hosseini-Asl, 2018)for calculating the encoder of each slot-value pair, the user utterance, and the previous system actions. However, we usean additional encoder for previous user utteranceto extract the historical and context.

The encoder model is used to encode the previous user utterance(H_{per}, C_{per}), the current userutterance (H_{cur}, C_{cur}), the previous system actions(H_a, C_a)for each of the system acts, and thecandidate (H_v, C_v)slot-value.As shown in Figure (1), a slot-embedding vector is used for k^{th} slot for context extraction.

To calculate representation H^k for eachslot k^{th} as shown in equation (1), we concatenated the slot embedding s_k with input sequence X, i.e., current user utterance, previous user utterance, or previous system actions, as input to the encoder, where concatenation is denoted as $f(X, s_k)$.

$$H^k = \mathrm{BiLSTM}(f(X, s_k)) \in \mathbb{R}^{n\times d_r} \quad (1)$$

where d_r is the dimension of the LSTM state. Then we calculate the attention scorea_i^k of the slot for each tokenhidden representationH_i^k as shown in equation (2), by concatenating them to the slot embedding $s_k$and transitory to a linear layer, then applying a softmaxp^k in equation (3) to normalize the distribution. In equation (4) compute similarly of the context c^k.

$$a_i^k = Wf(H_i^k, s_k) + b \in \mathbb{R} \quad (2)$$
$$p^k = \mathrm{softmax}(a^k) \in \mathbb{R}^n \quad (3)$$
$$c^k = \sum_i p_i^k H_i^k \in \mathbb{R}^{d_r} \quad (4)$$

Each of the fourencoders in the encoder model,as shownin Figure (1), can be representedas follows:

- U, $s_k$as inputs and H_u^k, c_u^k as outputs, where U denotes word embeddings of theuser utterance.
- P, s_k as inputs and H_p^k, p as outputs, where P denotes word embeddings of theprevious user utterance.
- A_j, s_k as inputs and H_{aj}^k, C_{aj}^k as outputs, where $A_j{}^{th}$is the previous system action.
- V, s_k as inputs and H_v^k, c_v^k as outputs, where V denotes current slot-value pair.

3.3. Scoring Model

We followed the proposed approach in GLAD (Zhong et al., 2018)for figuringthe score of each slot-value pair, in the current and previous user utterance and previous system actions.However, we usethe additional score toimprovecontext representation and distribution over dialogue history.

The scores model is used to compute slot k for its values to determine the value of the slot,which is mentioned by the users. Therefore, this was done usingfive sources. The firstscore$a_{cur_i}^k$ as shown in equation (5) is the current user utterance H_{cur}, taking into account the slot-value pair being considered c_v and using the resulting attention context q_{cur}equation (6) to score the slot-value pair.

$$a_{cur_i}^k = \mathrm{softmax}(\,(H_{cur_i}^k)^T c_v^k\,) \in \mathbb{R}^m \quad (5)$$
$$q_{cur}^k = \sum_i a_{cur_i}^k H_{cur_i}^k\,) \in \mathbb{R}^{d_r} \quad (6)$$
$$y_{cur}^k = Wq_{cur}^k + b \in \mathbb{R} \quad (7)$$

where mindicates a number of words in the inputsequence. The score y_{cur}^k equation (7) denotes the predicted valuesof the user utterance.

The secondscore$a^k_{pre_i}$as shown in equation (8) is similar to the first score, but uses previous user utterance H_{pre}instead of the current user utterance, taking into account the slot-value pair being considered c_v and using the resulting attention context q^k_{pre} equation (9) to score the slot-value pair.The score y^k_{pre} equation (10) denotes the predicted valuesof the previous utterance.

$$a^k_{pre_i} = softmax(\left(H^k_{pre_i}\right)^T c^k_v) \in \mathbb{R}^m \quad (8)$$

$$q^k_{pre} = \sum_i a^k_{pre_i} H^k_{pre_i}) \in \mathbb{R}^{d_r} \quad (9)$$

$$y^k_{pre} = Wq^k_{pre} + b \in \mathbb{R} \quad (10)$$

Then the predicted values of both current and previous user utterances are added as shown in the following equation (11):

$$y^k_u = y^k_{pre} + y^k_{cur} \quad (11)$$

Similarly, this is used to determine the mentionedprevious system actionsin the current or previous user utterance separatelyto reach sufficient information about user utterance when this is notinformative. The third score$a^k_{acur_i}$ as shown in equation (12), the context of current user utterance C_{cur} over the previous action representations $C_a = [C_{a1} \cdot \cdot \cdot C_{al}]$. Here, l is the number of previous system actions.Then we use the similarity between the attention context q_{acur}equation (13) and the slot-value pair c_v to score the slot-value pair.The score y^k_{acur} equation (14) denotes the predicted valuesof theprevious system actionsin the current user utterance separately.

$$a^k_{acur_i} = softmax(\left(C^k_{cur_i}\right)^T c^k_v) \in \mathbb{R}^{l+1} \quad (12)$$

$$q^k_{acur} = \sum_i a^k_{acur_i} C^k_{cur_i}) \in \mathbb{R}^{d_r} \quad (13)$$

$$y^k_{acur} = Wq^k_{acur} + b \in \mathbb{R} \quad (14)$$

The fourth source $a^k_{apre_i}$ as shown in equation (15),similar to the third score, but uses the context of previous user utterance C_{pre} inserted into the context of a current user utterance.

$$a^k_{apre_i} = softmax(\left(C^k_{pre_i}\right)^T c^k_v) \in \mathbb{R}^{l+1} \quad (15)$$

$$q^k_{apre} = \sum_i a^k_{apre_i} C^k_{pre_i}) \in \mathbb{R}^{d_r} \quad (16)$$

$$y^k_{apre} = Wq^k_{apre} + b \in \mathbb{R} \quad (17)$$

Then the predicted values of both the context of the current and previous user utterances are added:

$$y^k_a = y^k_{apre} + y^k_{acur} \quad (18)$$

In the last score,we followed the proposed approach in (Sharma, Choubey, & Huang, 2019)todetermine the relevance of the slot-value pairin the current turn. We concatenate the context of both current user utterance C^k_{cur} and previous user utterance C^k_{pre} and use a sigmoid activation of the linear layer to compute the score.

$$fc = W_{fc}\left(C^k_{cur} \oplus C^k_{pre}\right) + b_{fc} \quad (19)$$

$$\alpha = \sigma(W_\alpha \, tanh(fc) + b_\alpha) \quad (20)$$

Then wecompute context summaries $l_u$of attention from $C_v$overH_{cur} and $l_s$of attention from C_v over H_{per}.

$$l^k_u = Q(H^k_{cur}, c^k_v) \quad (21)$$

$$l^k_s = Q(H^k_{per}, c^k_v) \quad (22)$$

To compute the additional score y_f that establishesthe probability of the candidate slot-value based on both the current and previous user utterances and the previous system utterance, we use:

$$y^k_f = \alpha l^k_u + (1 - \alpha) l^k_s \quad (23)$$

Finally, we add the weight of all scores of slotk, i.e.,y^k_u, y^k_a, andy^k_f, which are normalized by the sigmoid function:

$$Y = \sigma(y^k_u + wy^k_a + y^k_f) \in \mathbb{R} \quad (24)$$

where w is a learned parameter.

IV. EXPERIMENT

Wizard of Oz (WoZ) is a singledomain of a restaurant reservation dataset (Wen et al., 2016)for dialogue tracking models. The emotion dataset consists of 600 dialogues for training,200 for evaluation, and 400 for testing. Each dialogue has an average of eight turns, where each turn contains a system utterance transcript, user utterance transcript, turn label, and dialogue state. The ontology consists of three different informable slottypes: food with 72 values, an area with 7 values, price range with 4 values, and requestswith 7 different slottypes like phone number andaddress.To evaluatethe metric of the joint goal tracking accuracy, we followed(Zhong et al., 2018)accumulation of turn goals.

4.1. Implementation Details

The pretrained GLoVeword embedding (Jeffrey Pennington, Richard Socher, 2014) is used, concatenated with character n-gram embeddings (Hashimotoet al., 2016), which are kept fixed during the training.We set the number of units inbi-LSTMs using 200 hidden dimensions and use the ADAM optimizer (Kingma & Ba, 2014)to train the models with the initial learning rate of 0.001. We set the dropout rate(Srivastavaet al., 2014)to 0.2 for the embedding layer. The maximum of epochs is set to 50 to train the models with a batch size of 50.

4.2. Results and Analysis

The previous user utterance in DST is used to gain necessary details to enhance the probability of distribution over all candidate slot-value pairs and action acts. Using one model to train and test that simplifies the whole process and gives the advantage of speed. The encoder side of the model is only calculated one time for the two tasks. We compute thescoresseparately of the previous user utterance and current user utterance as emotion in a section scoring model over all candidate slot-value pairs and action actstoimprovecontext representation and distribution over dialogue history, illustrated in (the scoring models of) Figure 1. Moreover,the fusion of the context of both previous user utterance and current user utterance then captures the distribution over all candidate slot-value pairs, illustrated in (the additional score of) Figure 1.The proposed model achieved outperforming results compared with all of the state-of-the-art approaches in the joint goal task, but that does not include the request turn task.

Table1:Comparison of our model to previously publishedWoZ restaurant reservation dataset

Model	Joint Goal	Turn Request
Neural Belief Tracker (NBT)—DNN	84.4%	91.2%
Neural Belief Tracker (NBT)—CNN	84.2%	91.6%
Global-Locally Self-Attentive (GLDA)	88.1%	97.1%
GloballyConditioned Encoder (GCE)	88.5%	**97.38%**
GCE **+** previous **utterance**	88.5**%**	96.98%
GCE **+** previous **utterance + additional score**	**89.3%**	96.98%

Table 1 shows a comparison of the performance of our proposed models with previous state-of-the-art models.The NBT(Mrkšić et al., 2016) applies CNN withpretraining word embeddings to representationlearning instead of delexicalized (Hendersonet al., 2014), which uses generic tags in its place slots and values in the utterance. In GLAD(Zhong et al., 2018)architecture, considered utterance, previous system action, and all slot values are separate in encoders. A similar architecture is used for all encoders.The authors used two scores model, each calculating the similarity of slot-value pairs to the user utterance representation or previous system action.

In the GCE (Nouri & Hosseini-Asl, 2018)model, the authors followeda similar approach to GLAD. However, the encoder was modified by removing the inefficient recurrent layers and self-attention layers to improve the latency and speed of inference.

The WoZ dataset was used for evaluatingour proposed model, implementing two attempted variants. The first was by using previous user utterance as an additional encoder in the dialogue using similar architecture to (Nouri & Hosseini-Asl, 2018), with the accuraciessimilar to the GCE model. The second was by using the additional score that combines the context of previous user utterance and current user utterance with all candidate slot-value pairs.Increasing current user utterance with appropriate information from the previous user utterance and previous system utterance further improves the joint goal accuracy, by 0.8%,and request accuracy, as is obvious from the results in Table 1.

V. CONCLUSION

The aim of DST is to determine two goals: the current state of each turn in dialogue slot-value pairs and the user's dialogue act to summarize all of the user's goals.In this paper, we figured the distribution of the slot-value pairs and dialogue acts in two separate ways to deal with the context of user utterance and by merging to provide the model with the necessary details of the information.Compared with all of the state-of-the-art approaches,the proposed model achieved outperforming results in the joint goal accuracy by 0.8% with the WoZdataset,whichused for evaluatingour proposed model.Our model uses a smaller number of learnable parameters that are added due to using relevant context for encoding current and previous user utterances. We also get high variance in the joint goal accuracy because the joint goal is calculated by stacking turn goals, and errors in predicting a turn goal are relatively fewer.

REFERENCES

[1]. Abhinav Rastogi , Dilek Hakkani-T ¨ ur, L. H. (2017). SCALABLE MULTI-DOMAIN DIALOGUE STATE TRACKING, 561–568.

[2]. Bohus, D., & Rudnicky, A. (2006). K hypotheses + other" belief updating model. Proceedings of AAAI Workshop on Statistical and Empirical Approaches to Spoken Dialogue Systems, 13–18. Retrieved from https://sites.google.com/site/yiqunhu/Home/distance-metric-learning

[3]. Casanueva, I., Budzianowski, P., Su, P.-H., Mrkšić, N., Wen, T.-H., Ultes, S., … Gašić, M. (2017). A Benchmarking Environment for Reinforcement Learning Based Task Oriented Dialogue Management. CoRR, abs/1711.1. Retrieved from http://arxiv.org/abs/1711.11023

[4]. Hashimoto, K., Xiong, C., Tsuruoka, Y., & Socher, R. (2016). A Joint Many-Task Model: Growing a Neural Network for Multiple NLP Tasks. Retrieved from http://arxiv.org/abs/1611.01587

[5]. Henderson, M., Thomson, B., & Young, S. (2013). Deep neural network approach for the dialog state tracking challenge. SIGDIAL 2013 - 14th Annual Meeting of the Special Interest Group on Discourse and Dialogue, Proceedings of the Conference, 467–471.

[6]. Henderson, M., Thomson, B., & Young, S. (2014a). Robust dialog state tracking using delexicalised recurrent neural networks and unsupervised adaptation. 2014 IEEE Workshop on Spoken Language Technology, SLT 2014 - Proceedings, 360–365. https://doi.org/10.1109/SLT.2014.7078601

[7]. Henderson, M., Thomson, B., & Young, S. (2014b). 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) (pp. 292–299). https://doi.org/10.3115/v1/W14-4340

[8]. Higashinaka, R., Nakano, M., & Aikawa, K. (2003). Corpus-based discourse understanding in spoken dialogue systems. Proceedings of 41st Annual Meeting on Association for Computational Linguistics, 240–247. https://doi.org/10.3115/1075096.1075127

[9]. J. Williams, A. Raux, D. Ramachandran, and A. B. (2017). The Dialog State Tracking Challenge with Bayesian Approach. Proceedings Of14th Annual Meeting of the Special Interest Group on Discourse and Dialogue, (August), 404–413. Retrieved from http://arxiv.org/abs/1702.06199

[10]. Jeffrey Pennington, Richard Socher, C. D. M. (2014). GloVe: Global Vectors for Word Representation. In Proceedings of the 2014 Con- Ference on Empirical Methods in Natural Language Processing (EMNLP), 1532–1543. https://doi.org/10.3115/v1/D14-1162

[11]. K. Sun, L. Chen, S. Zhu, and K. Y. (2014). A GENERALIZED RULE BASED TRACKER FOR DIALOGUE STATE TRACKING. Proceedings of Spoken Language Technology Workshop, 330–335.

[12]. Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization, 1–15. Retrieved from http://arxiv.org/abs/1412.6980

[13]. Lin, Z., Feng, M., Santos, C. N. dos, Yu, M., Xiang, B., Zhou, B., & Bengio, Y. (2017). A Structured Self-attentive Sentence Embedding, 1–15. Retrieved from http://arxiv.org/abs/1703.03130

[14]. Ma, Y., Raux, A., Ramachandran, D., & Gupta, R. (2012). Landmark-based location belief tracking in a spoken dialog system. SIGDIAL 2012 - 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue, Proceedings of the Conference, (July), 169–178.

[15]. Mrkšić, N., Séaghdha, D. O., Thomson, B., Gašić, M., Su, P. H., Vandyke, D., … Young, S. (2015). Multi-domain dialog state tracking using recurrent neural networks. ACL-IJCNLP 2015 - 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing, Proceedings of the Conference, 2, 794–799.

[16]. Mrkšić, N., Séaghdha, D. Ó., Wen, T.-H., Thomson, B., & Young, S. (2016). Neural Belief Tracker: Data-Driven Dialogue State Tracking. https://doi.org/10.18653/v1/P17-1163

[17]. Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and R. S. 2014. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfittin. Journal of Machine Learning Research, 1929–1958. https://doi.org/10.1016/0370-2693(93)90272-J

[18]. Nouri, E., & Hosseini-Asl, E. (2018). Toward Scalable Neural Dialogue State Tracking Model. Retrieved from http://arxiv.org/abs/1812.00899

[19]. Perez, J., & Liu, F. (2016). Dialog state tracking, a machine reading approach using Memory Network. Retrieved from http://arxiv.org/abs/1606.04052

[20]. Sepp, H., & Jurgen, S. (1997). Long Short-Term Memory. Neural Computation, 9(8), 1735–1780.

[21]. Sharma, S., Choubey, P. K., & Huang, R. (2019). Improving Dialogue State Tracking by Discerning the Relevant Context. Retrieved from http://arxiv.org/abs/1904.02800

[22]. Thomson, B., & Young, S. (2010). Bayesian update of dialogue state: A POMDP framework for spoken dialogue systems. Computer Speech and Language, 24(4), 562–588. https://doi.org/10.1016/j.csl.2009.07.003

[23]. Traum, S. L. and D. R. (2000). Information state and dialogue management in the trindi dialogue move engine toolkit.

[24]. Wen, T.-H., Vandyke, D., Mrksic, N., Gasic, M., Rojas-Barahona, L. M., Su, P.-H., … Young, S. (2016). A Network-based End-to-End Trainable Task-oriented Dialogue System. https://doi.org/10.1017/S1351324905004031

[25]. Xu, P., & Hu, Q. (2018). An End-to-end Approach for Handling Unknown Slot Values in Dialogue State Tracking. https://doi.org/10.1016/j.poly.2007.04.027

[26]. Zhong, V., Xiong, C., & Socher, R. (2018). Global-Locally Self-Attentive Dialogue State Tracker. Retrieved from http://arxiv.org/abs/1805.09655

[27]. Zilka, L., & Jurcicek, F. (2016). Incremental LSTM-based dialog state tracker. 2015 IEEE Workshop on Automatic Speech Recognition and Understanding, ASRU 2015 - Proceedings, 1(L), 757–762. https://doi.org/10.1109/ASRU.2015.7404864

[28]. Alhussayni, K. H., Zamyatin, A., & Alshamery, S. E. (2018). Bidirectional RNN-based Attention Model for Jointly Intent Detection and Slot Filling. Jour of Adv Research in Dynamical & Control Systems, 10(11), 404–410.

Khaldoon H. Alhussayni "Dialogue State Tracking Accuracy Enhancement by Distinguishing Candidate Slot-Value Pairs"International Journal Of Engineering Research And Development, vol. 15, no. 3, 2019, pp 23-28