using the provided train/test sets. The slot error rate and intent error rate are both set at 5%. We further evaluate our end-to-end framework on our DX dataset that reserves original conversation data. We selected 423 dialogues for training and conducted inference on another 104 dialogues.

Evaluation Metrics. Same as (Wei et al. 2018), we use the rate of making the right diagnosis as dialogue accuracy. We also employ a newly metric, namely matching rate, to evaluate the effectiveness of symptoms dialogue system requests. A successful matching means systems ask a symptom that exists in user implicit symptoms, otherwise, it is a failure matching.

Implementation Details We implement the system on Pytorch. To train the DQN composed of a two-layer neural network, the ϵ of ϵ-greedy strategy is set to 0.1 for effective action space exploration and the γ in Bellman equation is 0.9. The initial buffer size D is 10000 and the batch size is 32. The learning rate is 0.01. We choose SGD as the optimizer and 100 simulation dialogues will add to experience replay pool at each epoch training. Generally, we train the models for about 300 epochs. The source code will be released together with our DX dataset.

Experimental Results

MZ dataset. We only train Dialogue Management for the limitation of MZ dataset and compare our method with several baselines, as shown in Table 3."SVM-em" means the SVM model trained with just explicit symptoms and "SVM-em&im" is the SVM model using both explicit and implicit symptoms. Basic DQN is the proposed framework of (Wei et al. 2018). The results of the above three baselines are provided by (Wei et al. 2018). "DQN+relation branch" means our proposed model without knowledge-routed graph branch and "DQN+knowledge branch" is the model without relation refinement branch. Observed from Table 3, the performance of SVM-em&im is higher than SVM-em, which indicates that the implicit symptoms would make a significant improvement. However, basic DQN (Wei et al. 2018) gets 6% loss of accuracy compared to SVM-ex&im because it fails to inquiry effective implicit symptoms. Notably, our KR-DS not only significantly beats basic DQN (8%) but also outperforms SVM-ex&im (2%), which shows that our method can inquiry implicit symptoms effectively and make a precise diagnosis, thanks to knowledge-routed graph reasoning and relational refinement.

DX dataset. We further evaluate the proposed end-to-end KR-DS through Deep Q-learning on our DX dataset. For comparison, we re-implement a baseline, which shares the identical NLU and NLG with KR-DS but employ a single Deep Q-network (Wei et al. 2018) as policy network. In addition, we apply a state-of-the-art end-to-end task-oriented dialogue system framework (Lei et al. 2018) to our task (Sequicity), which uses belief spans to store constraints and requests (which are symptoms and diseases in this task) for state tracking. As shown in Table 4, our method outperforms basic DQN (Wei et al. 2018) and state-of-art seq-to-seq (Lei et al. 2018) method in both task complete accuracy and symptom matching rate. Focusing on obtaining the largest positive reward, basic DQN often guesses the right disease results but inquiries some unreasonable and repeated symptoms during the dialogue due to no constraints for symptom and disease relation (prior knowledge). Our framework shows superiority not only in a higher accuracy but also higher matching rate, which indicates the symptoms acquired by KR-DS agent is more reasonable and as a consequence, it can make the more right diagnosis. Seq-to-seq frameworks (Sequicity) performs worse on this medical diagnosis task as they focus on the in-dialogue sentence transition while ignoring medical symptom connections to diagnosis.

Method	Infantile diarrhea	Dyspepsia	Upper respiratory infection	Bronchitis	Overall
SVM-ex	0.89	0.28	0.44	0.71	0.59
SVM-ex&im	0.91	0.34	**0.52**	0.93	0.71
Basic DQN (Wei et al. 2018)	-	-	-	-	0.65
DQN+relation branch*	0.87	0.31	0.42	0.86	0.68
DQN+relation branch	0.92	0.35	0.49	0.93	0.70
DQN+knowledge branch	0.88	0.31	0.44	0.89	0.68
Our KR-DS	**0.96**	**0.39**	0.50	**0.97**	**0.73**

Table 3: Performance comparison on the MZ dataset.

Method	Accuracy	Match rate	Ave turns
Basic DQN (Wei et al. 2018)	0.731	0.110	3.92
Sequicity (Lei et al. 2018)	0.285	0.246	3.40
Our KR-DS	**0.740**	**0.267**	3.36

Table 4: Performance comparisons with the state-of-the-art methods on DX dataset.

Ablation Studies

Component analysis. To verify the effects of the main components of our KR-DS, we further conducted a series of ablation studies on MZ dataset, as shown in Table 3. Here we mainly target at the following components in our framework: knowledge-routed graph branch and relation refinement branch. As is shown in Table 3, all these factors contribute to better performance of our method. Additionally, initializing relation matrix with conditional probability in the relational branch is better than random initialization ("DQN+relation*"), as the prior medical knowledge can guide the relation matrix learning.

Reward evaluation. Our reward is designed based on the maximum turn value L=22, 2*L for success and -L for failure and -1 for the penalty. -1 penalty will cause shorter dialogue turns by accumulating through the process of dialogue. We evaluate several reward functions considering the magnitude of reward by doing experiments as follows: we chose four group of rewards R1: +22, -11, -1; R2: +11, -6,

Reward	R1	R2	R1*	R2*
Accuracy	0.697	0.725	0.718	0.739

Table 5: Evaluation of reward magnitude on MZ dataset.