```
{
  "disease_tag": "Pediatric hand, foot and mouth disease",
  "request_slots": {
    "diseases": "UNK"
  },
  "self-report": " Hello doctor, my baby keeps cry at sleep since the 2nd this  month,
  and he's been feeling uncomfortable at eat next day, we checked that not so
  much rashes on hands and feet. On third day, he's got a little fever and tonight I
  found so many rashes on his leg and feet. What is wrong with my baby? "
  "implicit_symptoms": {
      "Apathetic": True,
      "Anorexia": True
  }
}
```


Figure 3: An example of a user goal.

to obtain action probabilities. Subsequently, we sum these three results as predicted action distributions a_t of KR-DQN under the current state s_t.

$$a_t = sigmoid(a_t^r) + sigmoid(a_t^f) + a_t^k \tag{5}$$

In order to prevent repeated request, we add symptoms filter to KR-DQN outputs. All components are trained with a well-designed reward (described at the beginning of this section) to encourage KR-DQN learn how to request effective symptoms and make a right diagnosis.

User Simulator

In order to train our end-to-end dialogue system, we apply a user simulator to sample user goals from the experimental dataset for automatically and naturally interacting with the dialogue system. Following (Schatzmann and Young 2009), our user simulator maintains a user goal G. As shown in Fig. 3, a user goal generally consists of four parts: *disease tag* for the disease that the user suffers, *self-report* for the original self-reports from patients, *implicit symptoms* for symptoms talked about between the patient and the doctor, and *request slots* for the disease slot that the user would request. When the agent requests a symptom during the course of the dialogue, the user will take one of the three actions including *True* for the positive symptom, *False* for the negative symptom, and *Not sure* for the symptom that is not mentioned in the user goal. The dialogue session will be terminated as successful if the agent informs a correct disease. On the contrary, the dialogue process will fail if the agent makes the wrong diagnosis or the dialogue turn reaches the maximum turn *T*.

Natural Language Generation

Given actions produced from Dialogue Management and User Simulator, a template-based natural language generation (template-NLG) is applied in our system to generate human-like sentences. As mentioned in NLU part, request and inform pairs are relatively simple. Previous dialogue systems (Li et al. 2017; Lei et al. 2018) have many possible request/inform patterns but each one of them has only one template. Varying from them, we design 4 to 5 templates for each action to diversify dialogues. As for medical terms used in the dialogues, analogous to NLU, we choose daily expressions corresponding to specific symptoms and diseases from our collected medical term list.

Disease	**Quantity**	**Symptoms**
Allergic rhinitis	102	24
Upper respiratory tract infection	122	23
Pneumonia	100	29
Children hand-foot-mouth disease	101	22
Pediatric diarrhea	102	33

Table 2: Statistics of dialogues and symptoms for diseases.

End-to-End Training With Deep Q-Learning

Following (Mnih et al. 2015), we employ Deep Q-Learning to train DM with fine-tuned NLU and template-base NLG. Two important DQN tricks (Van Hasselt, Guez, and Silver 2016), target network usage and experience replay are applied in our system. We use $Q(s_t, a_t|\theta)$ to denote the the expected discounted sum of rewards, after taking a action a_t under state s_t. Then according Bellman equation, the Q-value can be written into:

$$Q(s_t, a_t|\theta) = r_t + \gamma \text{max}_{a_{t+1}} Q^*(s_{t+1}, a_{t+1}|\theta') \tag{6}$$

θ' is the parameters of target network obtained from previous episode. γ is the discount rate. We use ϵ-greedy exploration at training phase for effective action space exploration, selecting a random action in probability ϵ. We store the agents experiences at each time-step in experience replay buffer, denoted as $e_t(s_t, a_t, r_t, s_{t+1})$. The buffer is flushed if current network performs better than all previous models.

Experiments

DX Medical Dialogue Dataset

We build a newly DX dataset for medical dialogue system, reserving the original self-reports and interaction utterances between doctors and patients. We collected data from a Chinese online health-care community (dxy.com) where users asking doctors for medical diagnosis or professional medical advice. We annotate five types of diseases, including *allergic rhinitis*, *upper respiratory infection*, *pneumonia*, *children hand-foot-mouth disease*, and *pediatric diarrhea*. We extract the symptoms that appear in self-reports and conversation and normalize them into 41 symptoms. Four annotators with medical background are invited to label the symptoms in both self-reports and raw conversations.Symptoms appearing in self-reports are regarded as explicit symptoms while the others are implicit symptoms. The diseases of each medical diagnosis conversation are labeled automatically by the website. There are 527 conversational data in total. 423 conversational data are selected as the training set 104 for testing. More detailed dataset statistics are shown in Table .

Experimental Setup

Datasets. (Wei et al. 2018) constructed a dataset by collecting data from Baidu Muzhi Doctor website, denoted as MZ dataset in this paper. The MZ dataset contains 710 user goals and 66 symptoms, covering 4 types of diseases. As the MZ dataset only contains user goal data, we just train the DM model with user simulator and error model controller