End-to-end Generative Hierarchical Latent Intention Model for Task Oriented Dialogue Systems

Abstract

Developing task-oriented dialogue systems that can help common users make decisions and acquire useful knowledge through natural language has been a hot spot in both academic and industrial communities. The existing dialogue systems often heavily demand a huge amount of labeled data for natural language understanding, dialogue state tracking or policy network, which makes it costly to scale to other domains. Moreover, they are not fully end-to-end trainable and fail to capture the intrinsic variability and stochasticity during natural conversations. In this paper, we propose a hierarchical discrete latent intention encoder-decoder model to address these problems. We use a hierarchical encoder-decoder to characterize hierarchical structures with complex dependencies between subsequences. To model stochasticity of natural dialogues, we employ discrete latent variable inferred from users utterance to guide the generation process, which can be further refined via reinforcement learning under the identical framework with heuristic reward functions. Furthermore, we can estimate parameters of the model by optimizing exact log-likelihood instead of variational lower bound to achieve better results on success rate with a little sacrifice in BLEU score. The experiments on corpus-based and human evaluation demonstrate that our model outperforms the state-of-the-art models in terms of BLEU and achieves comparable success rate.

Introduction

Task-oriented spoken dialog systems aim to help people accomplish domain specific tasks through natural language (Young et al. 2013). However, it is not easy to build such a system due to the complex dialogue states and variability of natural language. Many efforts have been devoted to design different frameworks to tackle this challenge. Recurrent neural networks (RNNs) have demonstrated excellent results on numbers of machine leaning tasks involving sequential structured output form, such as machine translation (Sutskever, Vinyals, and Le 2014; Bahdanau, Cho, and Bengio 2014), natural language generation (Wen et al. 2015; Kiddon, Zettlemoyer, and Choi 2016), and language modeling (Mikolov et al. 2010). Recently, Sordoni et al. propose a hierarchical recurrent encoder-decoder (HRED) to model the structure of utterances dominated by statistics of the language and the dependencies among utterances in dialogues. It thus can generate reasonable responses that are related to conversation topics and speaker goals (Sordoni et al. 2015). Zhao et al. propose a similar architecture integrating entity indexing and augment with the chitchat ability (Zhao et al. 2017). However, these works are discriminative models that are trained to learn a conditional output distribution over strings and they do not consider the sophisticated architectures and conditioning mechanisms used to ensure salience. Hence, they exhibit limited abilities in handling the intrinsic variability and stochasticity of natural dialogues.

In order to tackle the problem above, a hierarchical latent variable encoder-decoder model (VHRED) is proposed by (Serban et al. 2017) to explicitly model generative processes with multiple levels of variability. VHRED introduces a continuous high-dimensional latent variable attached to each dialogue utterance, which can be optimized by maximizing the variational lower bound on the log-likelihood. Sharing the similar motivation for modeling stochasticity of dialogues, (Wen et al. 2017) propose a latent intention dialogue model (LIDM) to learn complex distributions of communicative intentions. They use discrete latent random variables to represent dialogue intentions as the sequential decision-making center of a dialogue agent, based on which the appropriate responses can be generated. Beyond that, they apply the neural variational inference learning to estimate parameters of LIDM. Actually, VHRED introduces continuous latent variable which is not interpretable from data and it lacks the ability of reinforcement learning fine-tuning which is critical for better dialogue modeling (Wen et al. 2017). In contrast, LIDM is proposed to model the variability of dialogues and interpretability of latent intentions. It revises conversational strategy based on an external reward under the same framework. However, LIDM follows the pipeline architecture that is used in (Young et al. 2013), which is not end-to-end trainable. At the same time, the errors may propagate from one module to the next one, which makes it hard to detect the bottleneck of the whole system. In addition, LIDM demands a huge amount of labeled data about the domain ontology to train a dialogue state tracker which is a key component (Rojas-Barahona et al. 2017) in dialogue systems. Furthermore, it heavily depends on predefined domain slots (Budzianowski et al. 2017), for scaling to other domains.

Motivated by these shortcomings, we develop a hierarchical discrete latent intention encoder-decoder (HDLIED) that is an end-to-end trainable system equipped with the ability of modeling latent intentions inferred from user's utterances and conducting refinement via reinforcement learning. Specifically, HDLIED is a hierarchical encoder-decoder model for modeling both utterances and dialogue context, where a discrete latent variable represents the user's intention. Instead of applying the neural variational inference learning (NVIL) (Mnih and Gregor 2014) to optimize parameters in the model, we directly maximize the exact log-likelihood benefiting from the latent intentions that follows the multi-nomial distribution. The exact log-likelihood objective function in our model needs to sum up all possible latent intentions of d-dimension followed by multi-nominal distribution, which is easy to compute since the number of latent intentions is d. The results on corpus-based and human evaluation demonstrate that using NVIL to conduct optimization, our model can achieve comparable BLEU scores and success rates compared to the state-of-the-art method (Wen et al. 2017) that requires the pre-defined domain ontology and massive labeled data. Moreover, the performance can be further improved by applying the exact MLE.

Related Work

Modeling chat-based dialogue as a sequence-to-sequence learning has been explored in the deep learning community (Sutskever, Vinyals, and Le 2014; Serban et al. 2017). Vinyals et al. demonstrate a neural conversation model by using the seq2seq model on a huge amount of dialogue data (Vinyals and Le 2015). However, it is a shallow (flat) generation process that tries to model the variability or stochasticity of dialogues by sampling words for output. Hence, it shows inability to model the dialogue context. Several works have been proposed to tackle the problem, such as modeling dialogue context through the hierarchical encoder-decoder framework (Serban et al. 2017; Zhao et al. 2017), introducing continuous latent variables (Sordoni et al. 2015; Cao and Clark 2017), and applying discrete latent variables (Wen et al. 2017). Distinguished from that, we not only exploit the hierarchical encoder-decoder for dialogue context modeling but also augment it with discrete latent variables to explicitly model the hidden dialogue intentions. Note that, previous latent variable models optimize variational lower bound, such as variational auto-encoder (VAE) (Kingma and Welling 2015) and neural variational inference learning (NVIL) (Mnih and Gregor 2014). In contrast, we directly optimize log-likelihood function by summing over latent variables that follow the multi-nominal distribution, which can be calculated in parallel fashion without sacrificing training efficiency.

At the other end of the spectrum, goal-oriented dialogue systems typically adopt the POMDP framework (Young et al. 2013) and break the system into a pipeline of modules including natural language understanding (NLU) (Barr 2017), dialogue state tracking (DST) (Williams, Raux, and Henderson 2016; Mrksic et al. 2017), dialogue policy network (Young et al. 2013) and natural language generation (NLG) (Wen et al. 2015). These dialogue models need predefined dialogue actions for a specific task. Therefore, they perform limited ability of scaling to more complex tasks. Furthermore, errors may propagate from previous modules to next modules and they are not easy to be detected. There have been some works for designing partially end-to-end trainable models by jointing several modules in this pipeline. For example, Yang et al. propose to joint NLU and dialogue policy learning (Yang et al. 2017). In contrast, HDLIED is end-to-end trainable and infers all underlying dialogue intentions from data. Modeling of end-to-end goal-oriented dialogue systems has been studied in (Rojas-Barahona et al. 2017; Bordes and Weston 2017). However, these models are typically deterministic and rely on decoder supervision signals to fine-tune a large set of model parameters. Different from them, our model with latent variables can be trained without any labeled data to tune parameters in the model. Clearly, it is easy to extend to new domains or datasets.

It is common to combine different learning paradigms to bootstrap performance. For example, reinforcement learning has been a common paradigm for dialogue modeling (Gasic et al. 2013; Su et al. 2016). Williams et al. propose a framework combing supervised learning and reinforcement learning to improve performance (Williams, Asadi, and Zweig 2017). Similarly, HDLIED parameterizes the intention space via a discrete latent variable and enjoys the benefit of bootstrapping from signals that come from different learning paradigms. In this paper, we incorporate unsupervised learning with reinforcement learning. We can also employ supervised, semi-supervised learning under the same framework for future work.

Proposed Model

HDLIED is an end-to-end trainable system under the framework of HRED (Sordoni et al. 2015) for dialogue context modeling augmenting with discrete latent variables attached to each dialogue utterance. It comprises three basic modules: (1) sentence embedding, (2) dialogue history modeling, and (3) response generation as shown in Fig 1. Given a dialogue $\mathcal{D} = \{d_1, d_2, \ldots, d_n\}$, where $d_t = (\boldsymbol{u}_t, \boldsymbol{r}_t)$, $\boldsymbol{u}_t = (u_{t,1}, \ldots, u_{t,m_{\boldsymbol{u}_t}})$ is user's utterance containing $m_{\boldsymbol{u}_t}$ words, and $\boldsymbol{r}_t = (r_{t,1}, \ldots, r_{t,m_{\boldsymbol{r}_t}})$ is agent's response containing $m_{\boldsymbol{r}_t}$ words at the t-th turn. Let $\boldsymbol{x}_t$ be a 2-dim one-hot vector named as "db_vector" that represents whether there are items matching user's constraints in the database, which is optional in our model.

Model Architecture

In a dialogue, the encoder RNN maps each utterance to an utterance vector as a distributed representation $\boldsymbol{h}_t$. The utterance vector is the hidden state obtained after the last token of the utterance is processed. In our model, we use a single-layer bi-directional LSTM as the utterance encoder.

$$\boldsymbol{h}_t = biLSTM(\boldsymbol{u}_t) \tag{1}$$

The higher-level context RNN keeps track of past utterances by processing each utterance vector $\boldsymbol{h}_t$ iteratively. The hidden state of the context RNN $\boldsymbol{s}_t$ represents a summary of the dialogues up to turn t (including turn t).

$$\boldsymbol{s}_t = LSTM(\boldsymbol{s}_{t-1}, \boldsymbol{h}_t) \tag{2}$$

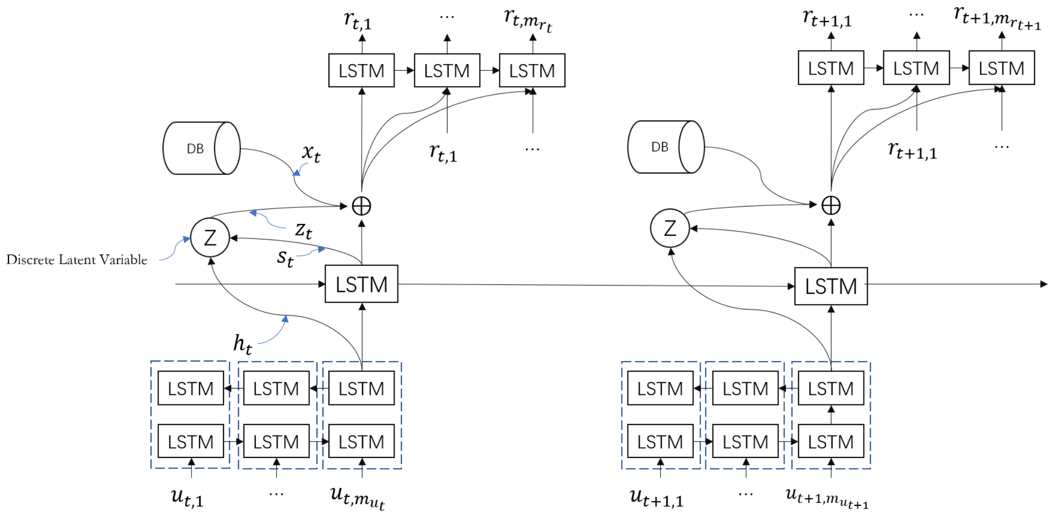

Figure 1: The architecture of HDLIED.

Conditioning on dialogue context $\boldsymbol{s}_t$ and current utterance vector $\boldsymbol{h}_t$, we use a single MLP layer to generate discrete latent variable for modeling user's intention to handle variability of dialogues. A latent intention $\boldsymbol{z}_t$ (or an action in the reinforcement learning literature) can then be sampled from the conditional distribution:

$$\boldsymbol{z}_t^l \sim \pi_{\Theta_2}(\boldsymbol{z}_t|\boldsymbol{s}_t, \boldsymbol{h}_t) \tag{3}$$

For response generation, we employ attention decoder (Bahdanau, Cho, and Bengio 2014) to generate better responses. Attention allows the decoder to look over every hidden state in the encoder and dynamically decide the importance of each hidden state at every decoding step. To be specific, let $\boldsymbol{h}_{t,m}$ and $\boldsymbol{c}_{t,m}$ denote the hidden state outputs of the encoder RNNs and the attention vector at turn t and time step m, respectively.

$$\boldsymbol{c}_{t,m} = \sum_i \alpha_{m,i}^t \boldsymbol{h}_{t,i} \tag{4}$$

where

$$\alpha_{m,i}^t = softmax(\boldsymbol{h}_{t,i}\boldsymbol{W}_a\boldsymbol{s}_{m-1}^t + \boldsymbol{b}_a) \tag{5}$$

and $\boldsymbol{s}_{m-1}^{\boldsymbol{t}}$ is the previous hidden state output by decoder RNNs at turn t and time $m-1$, $\{\boldsymbol{W}_a, \boldsymbol{b}_a\}$ are trainable parameters. The sampled intent $\boldsymbol{z}_t^l$, attention $\boldsymbol{c}_t$ and dialogue history $\boldsymbol{s}_t$ govern the generation of the response based on a conditional language generation model:

$$p(\boldsymbol{r}_t|\boldsymbol{d}_t, \boldsymbol{c}_t) = \prod_{l=1}^{m_{\boldsymbol{r}_t}} p_{\Theta_1}(r_{t,l+1}|r_{t,l}, \boldsymbol{s}_l^t, \boldsymbol{c}_{t,l}, \boldsymbol{d}_t) \tag{6}$$

where $\boldsymbol{d}_t = \boldsymbol{s}_t \oplus \boldsymbol{z}_t^l \oplus \boldsymbol{x}_t$ that $\oplus$ stands for vector concatenation.

Then the HDLIED can be formally written with its parameters $\Theta = \{\Theta_1, \Theta_2\}$:

$$p_\Theta(\boldsymbol{r}_t|\boldsymbol{s}_t, \boldsymbol{h}_t, \boldsymbol{c}_t, \boldsymbol{x}_t) = \sum_{\boldsymbol{z}_t} p_{\Theta_1}(\boldsymbol{r}_t|\boldsymbol{s}_t, \boldsymbol{c}_t, \boldsymbol{z}_t, \boldsymbol{x}_t)\pi_{\Theta_2}(\boldsymbol{z}_t|\boldsymbol{s}_t, \boldsymbol{h}_t) \tag{7}$$

Exact Maximum Log-likelihood Estimation

Equ. 7 is the objective function, which is similar to LIDM. However, LIDM applies NVIL (Mnih and Gregor 2014) to optimize parameters with inference network $q_\Phi(\boldsymbol{z}_t|\boldsymbol{s}_t, \boldsymbol{r}_t)$ to approximate true posterior $\pi_{\Theta_2}(\boldsymbol{z}_t|\boldsymbol{s}_t, \boldsymbol{h}_t)$. Notice that, since $\boldsymbol{z}_t$ follows multi-nominial distribution, we can actually sum over the all possible latent intentions to acquire Equ. 7. To be more specific, assuming that the latent intention $\boldsymbol{z}_t$ is a d-dim vector, there are only d different intentions. Hence summing over $\boldsymbol{z}_t$ is tractable and there is no need to use the variational lower bound to approixmate it. Executing exact maximum log-likelihood estimation (MLE) may not be feasible in practice due to its computational inefficiency especially when $\boldsymbol{z}_t$ is a high dimension vector. Thus we also apply NVIL to estimate parameters like LIDM. Much detailed information please refer to (Wen et al. 2017).

Reinforcement Learning

Using discrete latent variable as intention is able to control and refine the model's behaviour with operational experience. The learnt generative network $\pi_{\Theta_2}(\boldsymbol{z}_t|\boldsymbol{s}_t, \boldsymbol{h}_t)$ encodes the policy discovered from the underlying data distribution, but this is not necessarily the optimal for any specific task. Since $\pi_{\Theta_2}(\boldsymbol{z}_t|\boldsymbol{s}_t, \boldsymbol{h}_t)$ is a parameterised policy network itself, any policy gradient-based reinforcement learning (Williams 1992) can be used to fine tune policy against other objective functions that we are more interested in.

Based on the interested objective function J, we optimize it via the REINFORCE algorithm (Williams 1992).

$$\frac{\partial J}{\partial \Theta_2} \approx \frac{1}{M} \sum_{m=1}^{M} R_t^{(m)} \frac{\partial \log \pi_{\Theta_2}(\boldsymbol{z}_t^{(m)}|\boldsymbol{s}_t, \boldsymbol{h}_t)}{\partial \Theta_2} \tag{8}$$

For more information, please refer to the subsection **Training Details**.

Models	Success rate			BLEU		
	10	30	50	10	30	50
HRED	75.6%			**0.339**		
HRED,+db_vector	73.1%			0.330		
HDLIED	73.8%	76.3%	77.5%	0.290	0.283	0.292
HDLIED, +db_vector	71.9%	76.9%	81.9%	0.296	0.294	0.279
HDLIED, +RL	74.4%	76.3%	78.1%	0.287	0.283	0.288
HDLIED, +db_vector, +RL	73.8%	75.6%	**82.5%**	0.294	0.292	0.279

Table 1: Offline results of our HDLIED via **exact MLE** varying latent intention dimension from 10 to 50.

Models	Success rate				BLEU			
	30	50	70	100	30	50	70	100
VHRED	71.9%	73.8%	74.4%	72.5%	0.329	0.325	**0.337**	0.297
HDLIED	72.5%	72.5%	75.0%	73.8%	0.310	0.317	0.315	0.312
HDLIED, +db_vector	75.6%	75.6%	73.8%	78.1%	0.317	0.305	0.331	0.301
HDLIED, +RL	**84.5%**	73.1%	76.9%	80.0%	0.237	0.318	0.318	0.321
HDLIED, +db_vector, +RL	74.4%	75.6%	74.4%	78.1%	0.316	0.305	0.327	0.306

Table 2: Offline results of our HDLIED via **NVIL** (Mnih and Gregor 2014) varying latent intention dimensions from 30 to 100.

Experiments

Dataset and Metrics

We evaluate our proposed model by using the CamRest676 corpus collected by (Rojas-Barahona et al. 2017). The task is to assist users to find a restaurant in the CamBridge, UK area. There are three informable slots (food, pricerange, area) that users can use to constrain the search and six requestable slots (address, phone, postcode slots plus the three informable slots) that the user can ask for information about mentioned restaurant. There are 676 dialogues in the dataset (including both finished and unfinished dialogues) and approximately 2750 conversational turns in total. The database contains 99 unique restaurants. Like the previous works (Rojas-Barahona et al. 2017; Wen et al. 2017), the corpus was partitioned into training, validation and test sets in the ratio 3:1:1. The vocabulary size is 606 after preprocessing the original CamRest676 corpus[1].

We use task success rate (Su et al. 2015) and BLEU score (Papineni et al. 2002) to evaluate all models in the held-out test set. In order to assess the human evaluation performance, we evaluate HDLIED with optional blocks using exact maximum log-likelihood or variational lower bound for optimization, HRED, VHRED on our Web interface by recruiting subjects in company. Each judge was asked to follow a task and carried out a conversation with the machine. At the end of each conversation, the judges were asked to rate the models performance based on perceived comprehension ability, naturalness of responses and subjective success rate on a scale of 1 to 5 used in (Wen et al. 2017). For each model, we collected 50 dialogues and averaged the scores.

Published Models	Success rate	BLEU
NDM$^+$	76.1%	0.212
NDM+Att	79.0%	0.224
NDM+Att+SS	81.8%	0.240
LIDM*, I=50	66.9%	0.238
LIDM, I=70	61.0%	**0.246**
LIDM, I=100	63.2%	0.242
LIDM, I=50, +RL	82.4%	0.231
LIDM, I=70, +RL	81.6%	0.230
LIDM, I=100, +RL	**84.6%**	0.240

Table 3: Results of Published Models, where “+” are results of NDM from (Rojas-Barahona et al. 2017) with attention and self-supervised sub-task neurons, “$*$” are results of LIDM from (Wen et al. 2017) with various hyper-parameter settings.

Training Details

For all experiments, we set the word embedding size to 100. The hidden state size for both encoder RNNs and context RNN are 256. All encoders are bidirectional LSTMs. The forward and backward LSTMs both have 256 hidden units. The context RNN and decoder RNNs are GRU. The size of hidden states of context GRU is 256, and 512 for decoder GRU[2].

To make a direct comparison with LIDM, in which the latent intention sizes of model were set to 30, 50, 70 and 100 for variational inference learning used in (Wen et al. 2017) with **5** sampled intentions to calculate gradients and 10, 30, 50 for exact MLE. The trade-off factor λ is set to 0.1 (Higgins et al. 2017). All models[3] are trained at a learning rate of 0.0001. When trained with REINFORCE to further

[1] Similar to LIDM, all dialogues are pre-processed by delexicalisation (Henderson, Thomson, and Young 2014), where slot-value specific words are replaced with their corresponding generic tokens based on ontology.

[2] We concatenate hidden states of bidirectional encoder RNNs of the last step to initialize hidden states of decoder RNNs

[3] Except for training with RENIFORCE models

Metricss	Success	Naturalness	Comprehension	# of Turns
published models				
HRED	89.3%	3.98	4.01	4.60
VHRED	88.7%	3.70	3.91	4.42
NDM	91.5%	4.08	4.21	4.45
LIDM	92.0%	4.40	4.29	4.54
LIDM, +RL	93.0%	4.40	4.28	4.29
our models optimized via NVIL (Mnih and Gregor 2014)				
HDLIED	90.5%	4.01	4.10	4.40
HDLIED, +db_vector	90.9%	4.05	4.15	4.27
HDLIED, +RL	92.3%	4.30	4.19	4.50
HDLIED, +db_vector, +RL	91.5%	4.12	4.14	4.30
our models optimized via exact MLE				
HDLIED	91.5%	4.07	4.12	4.31
HDLIED, +db_vector	90.7%	4.09	4.19	4.29
HDLIED, +RL	91.3%	4.35	4.21	4.52
HDLIED, +db_vector, +RL	91.9%	4.14	4.23	4.45

Table 4: Results of human evaluation.

tune, we set the learning rate to 0.00001, and every mini-batch contains 32 training dialogues. We optimize all models in a end-to-end fashion using Adam (Kingma and Ba 2015) and tune (early stopping, hyper-parameters) on the held-out validation set. The drop rate in the context RNN and decoder RNNs with an adjustable number starts from 0.5 and linearly decreases to 0.0. When fine-tuned with REINFORCE, we only optimize the policy network by fixing the parameters of attention decoder. We introduce the reward function R_t which is described in LIDM, where constant η was set to 0.5.

$$R_t = \eta \cdot sBLEU(\boldsymbol{r}_t, \hat{\boldsymbol{r}}_t) + \begin{cases} 1, & \boldsymbol{r}_t \quad improves \\ -1, & \boldsymbol{r}_t \quad degrades \\ 0, & otherwise \end{cases} \qquad (9)$$

where $\boldsymbol{r}_t$ is the generated response, $\hat{\boldsymbol{r}}_t$ is the ground truth response, and sBLEU($\boldsymbol{r}_t$, $\hat{\boldsymbol{r}}_t$) is BLEU score of sentence. During testing, we greedily select latent intention and feed it into the decoder RNNs for response generation.

Experiments Results

Our model was assessed through both offline and online evaluations for exact MLE and NVIL (Mnih and Gregor 2014) with optional blocks including database vector and RL refinement. In Table 1 and Table 2, model using exact MLE achieves higher success even with lower dimension of latent intention than variational approximation. This phenomenon indicates that exact MLE could force model to generate slot-related token while model optimized via NVIL, will lose detail information due to the high variance when optimizing parameters of $q_\Phi(\boldsymbol{z}_t|\boldsymbol{s}_t, \boldsymbol{r}_t)$. But regrading BLEU, the latter achieves better results than the former. This may be due to the fact that variational lower bound of the dataset was optimised rather than task success rate using NVIL while exact MLE was optimized on task success rate. As for the impact of database vector and RL refinement, models with database vector improve the performance when optimizing variational lower bound but hurt it using exact log-likelihood objective function. Since we only sample 5 intentions to estimate parameters using NVIL, it may lose key information due to the limited sampled intentions and accuracy of variational approximation but database vector could compensate it by adding extra constraints. While for exact log-likelihood, it already maintains all intentions and extra constraints may be too strict to improve performance. Moreover, models using RL refinement optimized by exact MLE or NVIL show little improvements in all settings which is opposite to the results in LIDM (Wen et al. 2017). The reason may contribute to model architecture or definition of reward function since our model has a total different architecture compared to LIDM, thus making it hard to explore in the space of latent intention using similar reward function used in (Wen et al. 2017).

Compared to other existing models including (Rojas-Barahona et al. 2017) and (Wen et al. 2017) as presented in Table 3, our models optimized with exact log-likelihood or variational lower bound objective function could achieve comparable task success rate to LIDM (LIDM, I=100, +RL) and NDM (with attention and self-supervised sub-task neurons). It shows that our model, which is end-to-end trainable, can learn reasonable dialogue policy even without any labeled data. Augmenting with discrete latent intention, further improves the performances mainly on success rate compared to HRED which is a deterministic model, indicating that latent intentions could enlarge the model capacity for modeling uncertainty of natural dialogues. In addition, compared to VHRED, discrete latent variable may be more appropriate than continuous variable used in VHRED for modeling variability of dialogues for task oriented dialogue systems.

In Table 5 and Table 6, we list some samples from test dataset. When optimized using variational lower bound, HDLIED with RL prefers to generate slot related tokens to provide information of restaurants. Hence, it exhibits a much greedier behavior in achieving task success and shorter responses than HDLIED since the reward function Equ. 9 increases the weight on slot-related tokens. Moreover,

User	I need to find an expensive restauant that's in the south section of the city.
VHRED	what food are you looking for ?
HDLIED	what food would you like ?
HDLIED, +db_vector	what food are you looking for ?
HDLIED, +RL	the address is 82 Cherry Hinton Road Cherry Hinton.
HDLIED, +db_vector, +RL	what food are you looking for ?
User	no i don't care about the type of cuisine.
VHRED	the good luck is a expsneive restaurant in the south area.
HDLIED	the good luck chinese food takeaway serves chinese food and is in the expensive pricerange.
HDLIED, +db_vector	the good luck chinese food takeaway is a expensive restaurant in the south area. would you like their address ?
HDLIED, +RL	serve chinese food.
HDLIED, +db_vector, +RL	the good luck chinese food takeaway serves chinese food and is in the expensive pricerange.
User	what is their address ?
VHRED	82 Cherry Hinton Road Cherry Hinton .
HDLIED	82 Cherry Hinton Road Cherry Hinton .
HDLIED, +db_vector	82 Cherry Hinton Road Cherry Hinton .
HDLIED, +RL	82 Cherry Hinton Road Cherry Hinton .
HDLIED, +db_vector, +RL	82 Cherry Hinton Road Cherry Hinton .
User	what is the food they serve ?
VHRED	they serve chinese food .
HDLIED	they serve chinese food .
HDLIED, +db_vector	they serve chinese food .
HDLIED, +RL	they serve chinese food .
HDLIED, +db_vector, +RL	they serve chinese food .
User	nope , that's what i was looking for . thanks a lot !
VHRED	thank you .
HDLIED	thank you for using our system . good bye .
HDLIED, +db_vector	goodbye .
HDLIED, +RL	thank you .
HDLIED, +db_vector, +RL	thank you .

Table 5: Sampled dialogue results of VHRED and HDLIED using best models from Table 2 on test corpus set using **NVIL**.

the generated responses from HDLIED with different blocks (database vector and RL refinement) tend to generate similar responses. It may lie in that the distribution of learned intention may be peaked in several ones and the model prefers to sample one of them to generate responses that result in a lack of diversity. However, when optimized via exact MLE, the model not only tends to generate relevant information for user's constraints, but also provides extra information for user to request, which is more natural and comprehensible.

Conclusion

In conclusion, we propose a hierarchical discrete latent intention endocer-decoder model (HDLIED) for task oriented dialogue systems. To be more specific, HDLIED uses a hierarchical recurrent neural network structure (HRED) to model historical information of dialogue and augments it with discrete latent variables to represent latent intentions of users. Under our proposed framework, HDLIED can be trained in a fully end-to-end trainable unsupervised fashion with exact log-likelihood or variational lower bound. It achieves better scalability and easy to extend to other domains without costly data labeling. Augmenting with latent intentions leads to better generation performance due to its ability of modeling variability of natural dialogues. The experimental results on corpus and human evaluation show that HDLIED optimized with exact MLE or NVIL outperforms the state-of-the-art models in terms of BLEU scores and achieves comparable success rates. For future work, we will extend our model to multi-domain dialogue modeling and augment it with chitchat ability for better customer experience. We may also design new reward functions for reinforcement learning to bootstrap performance.

References

Bahdanau, D.; Cho, K.; and Bengio, Y. 2014. Neural machine translation by jointly learning to align and translate. *Computer Science*.

Barr, A. 2017. Natural language understanding. *AI Magazine* 1(1):5.

Bordes, A., and Weston, J. 2017. Learning end-to-end goal-oriented dialog. In *Proceedings of International Conference on Learning Representations (ICLR)*.

Budzianowski, P.; Ultes, S.; Su, P.; Mrksic, N.; Wen, T.; Casanueva, I.; Rojas-Barahona, L. M.; and Gasic, M. 2017. Sub-domain modelling for dialogue management with hierarchical reinforcement learning. In *Annual SIGdial Meeting on Discourse and Dialogue (SIGDIAL)*.

Cao, K., and Clark, S. 2017. Latent variable dialogue models and their diversity. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*, volume 2.

Gasic, M.; Breslin, C.; Henderson, M.; Kim, D.; Szummer, M.; Thomson, B.; Tsiakoulis, P.; and Young, S. J. 2013. On-

User	I need to find an expensive restauant that's in the south section of the city.
HRED	the good luck chinese food takeaway is a expensive restaurant in the south area. would you like their information ?
HDLIED	what food are you looking for ?
HDLIED, +db_vector	what food are you looking for ?
HDLIED, +RL	what food are you looking for ?
HDLIED, +db_vector+RL	what food are you looking for ?
User	no i don't care about the type of cuisine .
HRED	the good luck chinese food takeaway serves chinese food and is in the expensive pricerange.
HDLIED	there is a expensive chinese restaurant in the expensive pricerange. would you like the phone ?
HDLIED, +db_vector	there is The Good Luck . would you like the information ?
HDLIED, +RL	there is a expensive chinese restaurant in the expensive pricerange. would you like the phone ?
HDLIED, +db_vector+RL	there is The Good Luck . would you like the information ?
User	what is their address ?
HRED	82 Cherry Hinton Road Cherry Hinton .
HDLIED	The Good Luck is located at 82 Cherry Hinton Road Cherry Hinton. is there anything else i can help you with today ?
HDLIED, +db_vector	the address for The Good Luck is 82 Cherry Hinton Road Cherry Hinton. would you like their phone ?
HDLIED, +RL	there address is 82 Cherry Hinton Road Cherry Hinton , is there anything else i can help you with ?
HDLIED, +db_vector, +RL	there address is 82 Cherry Hinton Road Cherry Hinton , is there anything else i can help you with ?
User	what is the food they serve ?
HRED	they serve chinese food .
HDLIED	they serve chinese food .
HDLIED, +db_vector	they serve chinese food .
HDLIED, +RL	they serve chinese food .
HDLIED, +db_vector, +RL	they serve chinese food .
User	nope , that's what i was looking for . thanks a lot !
HRED	you're welcome .
HDLIED	thank you .
HDLIED, +db_vector	goodbye .
HDLIED, +RL	thank you .
HDLIED, +db_vector, +RL	goodbye .

Table 6: Sampled dialogue results of HRED and HDLIED using best models from Table 1 on test corpus set using **exact MLE**.

line policy optimisation of bayesian spoken dialogue systems via human interaction. In *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2013, Vancouver, BC, Canada, May 26-31, 2013*, 8367–8371.

Henderson, M.; Thomson, B.; and Young, S. 2014. Word-based dialog state tracking with recurrent neural networks. In *Meeting of the Special Interest Group on Discourse and Dialogue*, 292–299.

Higgins, I.; Matthey, L.; Pal, A.; Burgess, C.; Glorot, X.; Botvinick, M.; Mohamed, S.; and Lerchner, A. 2017. beta-vae: Learning basic visual concepts with a constrained variational framework. In *Proceedings of International Conference on Learning Representations (ICLR).*

Kiddon, C.; Zettlemoyer, L.; and Choi, Y. 2016. Globally coherent text generation with neural checklist models. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016*, 329–339.

Kingma, D., and Ba, J. 2015. Adam: A method for stochastic optimization. In *The International Conference on Learning Representations (ICLR).*

Kingma, D. P., and Welling, M. 2015. Auto-encoding variational bayes. In *Proceedings of International Conference on Learning Representations (ICLR).*

Mikolov, T.; Karafiát, M.; Burget, L.; Cernockỳ, J.; and Khudanpur, S. 2010. Recurrent neural network based language model. In *Proceedings of the 11th Annual Conference of the International Speech Communication Association (INTERSPEECH).*

Mnih, A., and Gregor, K. 2014. Neural variational inference and learning in belief networks. In *Proceedings of the 34th International Conference on Machine Learning (ICML).*

Mrksic, N.; Séaghdha, D. Ó.; Wen, T.; Thomson, B.; and Young, S. J. 2017. Neural belief tracker: Data-driven dialogue state tracking. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, 1777C1788.

Papineni, K.; Roukos, S.; Ward, T.; and Zhu, W. J. 2002. Bleu: a method for automatic evaluation of machine translation. In *Meeting on Association for Computational Linguistics*, 311–318.

Rojas-Barahona, L. M.; Gasic, M.; Mrksic, N.; Su, P.; Ultes, S.; Wen, T.; Young, S. J.; and Vandyke, D. 2017. A network-based end-to-end trainable task-oriented dialogue system. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, EACL*

2017, Valencia, Spain, April 3-7, 2017, Volume 1: Long Papers, 438–449.

Serban, I. V.; Sordoni, A.; Lowe, R.; Charlin, L.; Pineau, J.; Courville, A. C.; and Bengio, Y. 2017. A hierarchical latent variable encoder-decoder model for generating dialogues. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI)*, 3295–3301.

Sordoni, A.; Bengio, Y.; Vahabi, H.; Lioma, C.; Grue Simonsen, J.; and Nie, J.-Y. 2015. A hierarchical recurrent encoder-decoder for generative context-aware query suggestion. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, 553–562. ACM.

Su, P. H.; Vandyke, D.; Gasic, M.; Kim, D.; Mrksic, N.; Wen, T. H.; and Young, S. 2015. Learning from real users: Rating dialogue success with neural networks for reinforcement learning in spoken dialogue systems. In *Proceedings of the 11th Annual Conference of the International Speech Communication Association (INTERSPEECH)*.

Su, P.; Gasic, M.; Mrksic, N.; Rojas-Barahona, L. M.; Ultes, S.; Vandyke, D.; Wen, T.; and Young, S. J. 2016. On-line active reward learning for policy optimisation in spoken dialogue systems. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, 2431C2441.

Sutskever, I.; Vinyals, O.; and Le, Q. V. 2014. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, 3104–3112.

Vinyals, O., and Le, Q. V. 2015. A neural conversational model. In *Proceedings of the International Conference on Machine Learning, Deep Learning Workshop*.

Wen, T.; Gasic, M.; Mrksic, N.; Su, P.; Vandyke, D.; and Young, S. J. 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*, 1711C1721.

Wen, T.-H.; Miao, Y.; Blunsom, P.; and Young, S. 2017. Latent intention dialogue models. In Precup, D., and Teh, Y. W., eds., *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, 3732–3741. International Convention Centre, Sydney, Australia: PMLR.

Williams, J. D.; Asadi, K.; and Zweig, G. 2017. Hybrid code networks: practical and efficient end-to-end dialog control with supervised and reinforcement learning. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, 665C677.

Williams, J.; Raux, A.; and Henderson, M. 2016. The dialog state tracking challenge series: A review. *Dialogue & Discourse* 7(3):4–33.

Williams, R. J. 1992. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning* 8(3-4):229–256.

Yang, X.; Chen, Y.; Hakkani-Tür, D. Z.; Crook, P.; Li, X.; Gao, J.; and Deng, L. 2017. End-to-end joint learning of natural language understanding and dialogue manager. In *The 42nd IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, volume abs/1612.00913.

Young, S.; Gašić, M.; Thomson, B.; and Williams, J. D. 2013. Pomdp-based statistical spoken dialog systems: A review. *Proceedings of the IEEE* 101(5):1160–1179.

Zhao, T.; Lu, A.; Lee, K.; and Eskénazi, M. 2017. Generative encoder-decoder models for task-oriented spoken dialog systems with chatting capability. In *18th Annual SIGdial Meeting on Discourse and Dialogue (SIGDIAL)*.