Converse, Focus and Guess - Towards Multi-Document Driven Dialogue

Han Liu[1], **Caixia Yuan**[1*], **Xiaojie Wang**[1],
Yushu Yang[2], **Huixing Jiang**[2], **Zhongyuan Wang**[2]
[1]Beijing University of Posts and Telecommunications, [2]Meituan-Dianping Group
{liuhan,yuancx,xjwang}@bupt.edu.cn, {yangyushu, jianghuixing, wangzhongyuan02}@meituan.com

Abstract

We propose a novel task, **M**ultiple **D**ocuments-**D**riven **D**ialogue (**MD3**), in which an agent can guess the target document that the user is interested in by leading a dialogue. To benchmark progress, we introduce a new dataset of GuessMovie, which contains 16,881 documents, each describing a movie, and associated 13,434 dialogues. Further, we propose the MD3 model. Keeping guessing the target document in mind, it converses with the user conditioned on both document engagement and user feedback. In order to incorporate large-scale external documents into the dialogue, it pretrains a document representation which is sensitive to attributes it talks about an object. Then it tracks dialogue state by detecting evolvement of document belief and attribute belief, and finally optimizes dialogue policy in principle of entropy decreasing and reward increasing, which is expected to successfully guess the user's target in a minimum number of turns. Experiments show that our method significantly outperforms several strong baseline methods and is very close to human's performance.

Introduction

The recent progress with human-machine dialogue techniques enable conversational agents to be extensively applied in customer service, information retrieval, personal assistance and so on. In order to assist the user to accomplish specific tasks, the agent must necessarily query external knowledge. Several works have focused on incorporating structured knowledge base (KB) into dialogues (Dhingra et al. 2017; Madotto, Wu, and Fung 2018; Wu, Socher, and Xiong 2019) through KB lookup. Although these efforts scales nicely to huge knowledge base, many real-world task-oriented dialogues involve in referring to a great number of documents (such as manuals, instruction booklets, and other informational documents). Since the complexity of document understanding, developing dialogues with many grounding articles is far from a trial task.

In this paper, we consider in particular the problem of multi-document driven dialogue (MD3), where the agent leads a dialogue with the user with a particular conversation goal and the engagement of multiple documents. To this end, we propose a MD3 game—*GuessMovie*. In the game, the user selects a movie at the beginning of dialogue, which is unknown to the agent. The agent is provided with a set of candidate documents, each describing a movie, and tries to guess which movie the user selects by asking a series questions (e.g. "*Who is the director of the movie?*" or "*When is it released?*"). The user informs the agent his answer or says "*unknown*". The agent's goal is to guess the target movie in the shortest dialogue turns. It assumes that only the agent has access to the documents, and at least one document describing the target movie. Figure 1 shows an example.

Two key challenges arising from MD3 task is: (1) how to efficiently encode and incorporate a large scale of long unstructured documents in a dialogue, and (2) how to learn optimal dialogue policy to efficiently fulfill dialogue task with smooth engagement of external documents.

To address the first challenge, we propose an factor aware document embedding, which assumes that a text is typically composed of several key factors it wants to narrate. For example, a text describing a movie mainly consists of attributes like "*directed by*", "*release year*", "*genre*", and so on. For each attribute, attribute aware document embedding is trained with the goal of correctly recognizing the document containing the given attribute and attribute value from a set of documents. The document is finally represented as the concatenation of all attribute aware document embeddings.

In order to manipulate the dialogue towards efficiently guessing the correct target document, the proposed MD3 model traces dialogue state by monitoring a document belief distribution and attribute belief distribution, both in global dialogue-level and temporal turn-level, which is viewed an explicit representation of dialogue dynamics leading to target movie. Besides, the model also calculates the uncertainty of each attribute through a document differentiated representation. In order to fulfill the dialogue goal within minimum dialogue turns, at each turn the policy model tends to discussing the attribute with the highest belief and highest uncertainty with *asking* the user. When the belief of a document is higher than a threshold, the agent will execute a *guess* action. A dialogue is deemed as ended up with a *guess* action.

Remarkably, the above dialogue state tracing and dialogue policy optimizing are jointly trained using reinforcement learning.

Our contributions are summarized as follows:

1. We propose a novel MD3 task, and release a new publicly-

*Corresponding author.

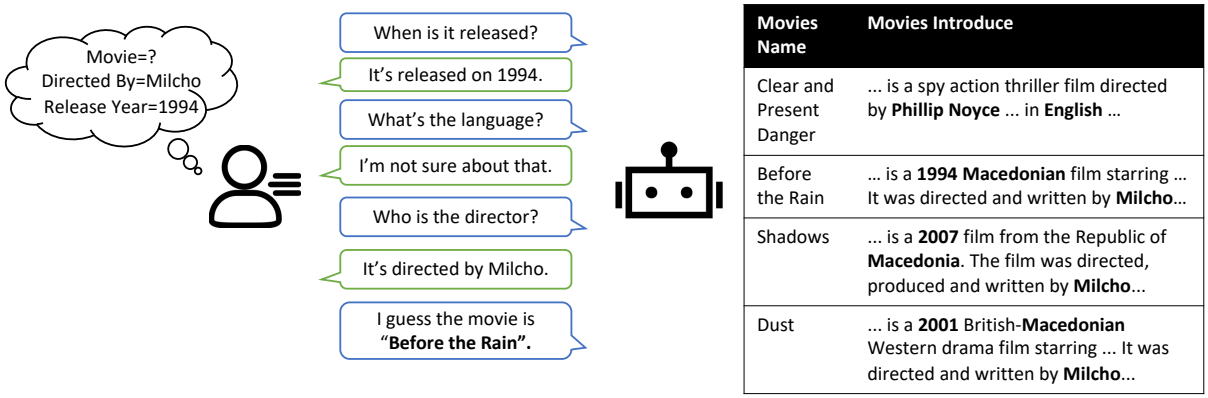


Figure 1: Sample dialogue from GuessMovie dataset. The user's target is "*Before the Rain*" and only knows its director is Milcho, and release year is *1994*. The goal of the agent is to correctly guess the target movie via asking minimum number of questions to the user. Here we only demonstrate four candidate movie documents due to space constraints (any number of documents could be provided in practice). Since candidate movies have the highest uncertainty (i.e. the largest information entropy) on "*release year*", the agent ask "*When is it released?*" to minimize the range of candidates. After the first turn, the last two movies "*Shadows*" and "*Dust*" can be excluded. Similarly, the agent ask "*in language*" and "*directed by*" in the second and third turn.

available benchmark dialogue corpus—GuessMovie, that we hope will help further work on document-driven task-oriented dialogue agents.

2. We introduce the MD3 model, a highly performant neural dialogue agent that is able to smoothly incorporate multiple documents through entangling document representation, document belief and attribute belief to the dynamics of the dialogue. As far as we know, this is the first study of task-oriented dialogue based on a large scale of documents.
3. The proposed document-aware dialogue policy achieve the maximum dialogue success rate in the minimum number of dialogue turns compared with several baseline models.

Related Work

The closest work to ours lies in the area of dialogue system incorporating unstructured knowledge. Ghazvininejad et al. (2018); Parthasarathi and Pineau (2018) use an Encoder-Decoder architecture where the decoder receives as input an encoding of both the context utterance and the external text knowledge. Dinan et al. (2018); Li et al. (2019) investigate extended Transformer architecture to handle the document knowledge in multi-turn dialogue. Reddy, Chen, and Manning (2019) use documents as knowledge sources for conversational Question-Answering. This line of work aims either at producing more contentful and diverse responses, or at extracting answers of user questions. The dialogue agent we build has a specific goal throughout the dialogue, from this point of view the dialogue system in this paper is task-oriented one. While these works mainly focusing on chatting about the content of a given document without a specific dialogue goal. Besides, our task differs from them in that our agent interact with a large-scale external documents, which poses new challenges for grounding dialogues.

Another line of related work is on "*Guess*"-style dialogues, among which *Q20 Game* (Burgener 2006; Zhao and Eskenazi 2016; Hu et al. 2018) is a typical object guessing game. In Q20, the agent guesses the target object within 20 turns of questions and answers. Each object is tied with a structured KB and the user is restricted to passively answer "*Yes, No or Unknown*". Dhingra et al. (2017) proposes a new task and method named KB-InfoBot, which can be regarded as an extension of Q20 Game. It aims to retrieval from the structured KB through a dialogue by a soft query operation. These works mainly focus on integrating structured knowledge into dialogue systems, while it requires a lot of work to build up, and is only limited to expressing precise facts. Documents as knowledge are much easier to obtain and provide a wide spectrum of knowledge, including factoid, event logics, subjective opinion, etc.

In the field of computer vision, both the ImageGuessing (Das et al. 2017) and GuessWhat (De Vries et al. 2017) try to guess a picture or an object through multi-turn dialogue, which greatly expands the range of dialogue applications (Pang and Wang 2020b,a). We use a similar approach to extend dialogue games into the document-driven ones. Different from the vision field, how to encode large-scale text information is a vital challenge.

Dataset: GuessMovie

We build a benchmark *GuessMovie* dataset for MD3 task on the base of the dataset WikiMovies (Miller et al. 2016). In WikiMovies, there is a large-scale document knowledge. Each is a movie introduction text, which is derived from the first paragraph of Wikipedia. In addition, it also contains a structured KB of the same movie collection, which is derived from MovieLens, OMDB and other datasets. There are totally 10 different attributes in movie KBs. But we select the common 6 ones since others are randomly discussed in the corresponding documents.

As for GuessMovie, we firstly align a document with a piece of structured knowledge. For a specific movie, some attribute might be missing from the text and the KB, and some

attribute might hold more than one values (e.g. a movie may have more than one starred actors.). Such cases are preserved in GuessMovie, which is expected to make the dialogue more diverse and realistic.

Then we create multi-turn dialogues for some documents. We design a dialogue simulator that interacts with the structured KB to generate dialogues. It consists of two agents playing the roles of the user and the system. Both agents interact with each other using a finite set of dialogue acts directing the dialogue.

Totally, GuessMovie is comprised up with 13,434 dialogues for 16,881 documents. The average length of documents is 107.66 words. More statistics are described in Table 1.

To build the GuessMovie dataset with dialogues, we use a user simulator introduced in the following section. The system agent is provided with a candidate KB, including the target knowledge and a set of other randomly selected knowledge. At the beginning, the system agent generates a dialogue act of "*asking*" an attribute (e.g. "*when is the movie released?*"). The probability of an attribute being chosen as "*asked attribute*" is proportional to the information entropy computed according to its distribution within the candidate KB. After a turn of "question-answering", the KBs insistent with the facts so far are filtered out. The dialogue continues until the system agent is confident with the target KB and executes a "*guess*" action. For natural language generation, we use several diversed natural language patterns that takes an attribute or an attribute-value pair as argument.

It's worth noting that we do not include any domain-specific constraints in both simulated agents. Although our examples use Wikipedia articles about movies, we see the same techniques being valid for other external documents such as manuals, instruction booklets, and other informational documents, as far as they can be loosely structed as several facets about an object.

Method

Figure 2 illustrates the overall architecture of the ***Multi-Document Driven Dialogue (MD3)*** model, including five parts: document representation, Natural Language Understanding (NLU), Dialogue State Tracking (DST), Policy Model (PM) and Natural Language Generation (NLG). This section introduces each part in details.

Task Definition Given a set of M documents $\{D_i\}_{i=1}^{M}$ as candidates, and a target document known only by the user (each of which corresponds to an unique object, i.e. a movie in our case), the agent outputs a consecutive of responses that ask an attribute (e.g., "*when is it released?*", "*Is it released in 1990?*"), or guesses a target document and ends the dialogue. The user can provide the answer to the questions, or answer like "*I don't know*".

AaDR for Document Representation

As for encoding large-scale documents along with optimizing dialogue agent is computationally burdening and inviable, we resort to pretrain the documents in advance. We argue that the pre-trained representation should be not only liable to the original meaning but also helpful for constructing a document-driven task-oriented dialogue agent.

To this end, we propose ***Attribute-aware Document Representaion (AaDR)***. We assume that each document talks about several attributes with index $\{1, ..., j, ..., L\}$, such as "*directed by*", "*release year*", "*in language*" and so on in movie scenario.

Inspired by Hierarchical Attention Networks (HAN) (Yang et al. 2016) and Hierarchical Label-Wise LSTM (Liu, Yuan, and Wang 2020), we introduce an *Attribute-aware HAN (AaHAN)* to encode each document, which seen attribute as label. Specifically, each attribute is used to index the corresponding parameters in hierarchical attention. This mechanism can capture different information for different attributes. Combined with the contrastive loss (Hadsell, Chopra, and LeCun 2006), an attribute-aware document representation is learned.

Pre-training Randomly sample a target document as T. For an attribute-value pair (a_j, v_{jk}), we sample a positive document C^+ which has the same value v_{jk} for attribute a_j as T and several negatives C^- which have different attributes values. These two parts are combined to a training sample $\{C_i\}_{i=1}^{R}$ for a_j. The training goal is to distinguish the positive document from all $\{C_i\}_{i=1}^{R}$.

As for the target T and a candidate C_i, we can calculate the corresponding document representation on a certain attribute a_j: $H_j^t \in \mathbb{R}^{2d}$ and $H_j^{c_i} \in \mathbb{R}^{2d}$. Further, calculate the similarity of two documents directly as follows.

$$H_j^t = \mathrm{AaHAN}^t(T, a_j) \tag{1}$$

$$H_j^{c_i} = \mathrm{AaHAN}^c(C_i, a_j) \tag{2}$$

It's trained using a negative log-likelihood loss function.

$$\mathop{\mathbb{E}}_{t,j,c^+,c^-}\left[-\log \frac{\exp\left((H_j^t)^T H_j^{c^+}\right)}{\exp\left((H_j^t)^T H_j^{c^+}\right) + \sum_{c^-}\exp\left((H_j^t)^T H_j^{c^-}\right)}\right] \tag{3}$$

Representation After pre-training, for each document D_i, we can obtain the representation $Q_{ij} \in \mathbb{R}^{4d}$ on attribute a_j, which is the concatenation of outputs of target encoder and candidate encoder.

$$H_{ij}^t = \mathrm{AaHAN}^t(D_i, a_j) \tag{4}$$

$$H_{ij}^c = \mathrm{AaHAN}^c(D_i, a_j) \tag{5}$$

$$Q_{ij} = [H_{ij}^t; H_{ij}^c] \tag{6}$$

After that, the final document representation $Q_i \in \mathbb{R}^{L\times 4d}$ is obtained by concatenating L attribute-aware representation.

DaLU for NLU

We propose ***Document-aware Langauge Understanding (DaLU)*** for NLU. In the previous turn, the agent's question is x^t, and the user's response is u^t, which are together concatenated into a long sentence and encoded using a BiGRU. Take the last hidden state as output $G^t \in \mathbb{R}^{2d}$.

Attr	directed by	release year	written by	starred actors	has genre	in language
Num	14853	16299	12712	13204	12118	3071
Ent	6187	103	10404	10180	23	96
Ave	1.05	1.00	1.48	2.48	1.27	1.10

Table 1: GuessMovie dataset statistics. *Num* denotes the number of documents containing it (with total 16,881 documents). *Ent* denotes the number of rare values. *Ave* denotes the average number of values that an attribute has.

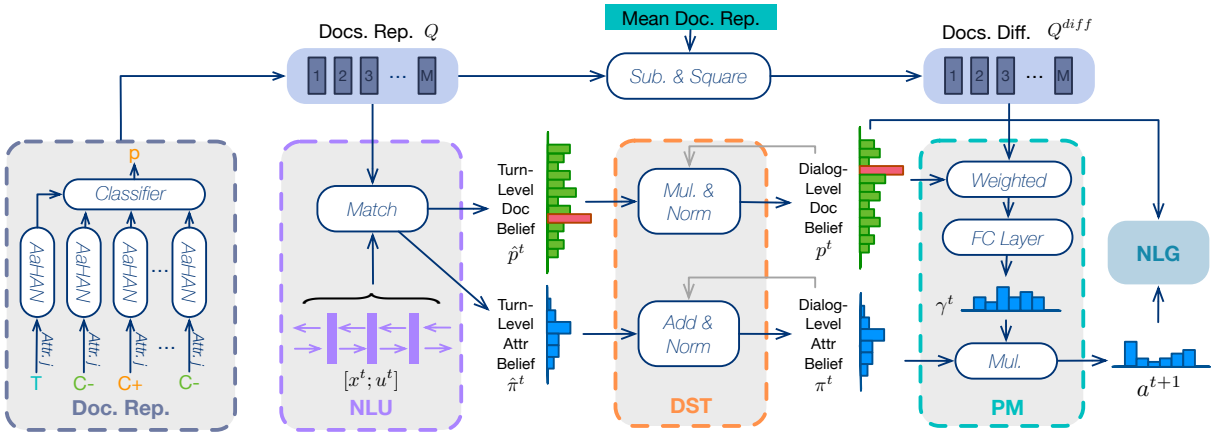


Figure 2: Overall architecture of MD3 model with AaDR for document representation, DaLU for NLU, DaST for DST, DaPO for PM and rule for NLG.

Therefore, the similarity $\hat{S}^t \in \mathbb{R}^{M\times L}$ between previous turn $G^t \in \mathbb{R}^{2d}$ and candidate documents $Q \in \mathbb{R}^{M\times L\times 4d}$ (concatenation of each candidate Q_i etc.) can be calculated directly by a bilinear method. It reflects the matching degree for each candidate Q.

$$\hat{S}^t = G^t W^s Q^T \tag{7}$$

where $W^s \in \mathbb{R}^{2d\times 4d}$ is a trainable parameter.

In addition, we further calculate two distributions: (1) the attribute type $\tilde{\pi}^t \in \mathbb{R}^L$. (2) a flag $\alpha^t \in \mathbb{R}^1$ indicating whether the response is "*unknown*". The closer the value is to 1, the less valid information is included in this turn.

$$\tilde{\pi}^t = \text{softmax}(W^{attr} G^t) \tag{8}$$

$$\alpha^t = \text{sigmoid}(W^{unk} G^t) \tag{9}$$

where $W^{attr} \in \mathbb{R}^{L\times 2d}$ and $W^{unk} \in \mathbb{R}^{1\times 2d}$ are trainable parameters.

When previous response is "*unknown*", the selected probability of each candidate is equal on turn level. Therefore, we expand the similarity $\hat{S}^t$ on attribute dimension, concatenating a vector whose similarity is all 1, and get $S^t \in \mathbb{R}^{M\times(L+1)}$. For the attribute type $\tilde{\pi}^t$, we also expand the attribute dimension by fusing the $\tilde{\pi}^t$ and α^t, and get $\beta^t \in \mathbb{R}^{(L+1)}$.

$$S^t = [\hat{S}^t; \mathbb{1}] \tag{10}$$

$$\beta^t = [\tilde{\pi}^t(1-\alpha^t); \alpha^t] \tag{11}$$

Further the Turn-Level Doc [1] Belief $\hat{p}^t \in \mathbb{R}^M$ is obtained, which is the probability of each candidate document being selected at current turn, and also the Turn-Level Attr Belief $\hat{\pi}^t \in \mathbb{R}^L$.

$$\hat{p}^t = \text{softmax}(S^t \beta^t) \tag{12}$$

$$\hat{\pi}^t = \alpha^t \tilde{\pi}^t \tag{13}$$

DaST for DST

We propose ***D**ocument-**a**ware **S**tate **T**racking (DaST)* for DST. The *dialogue state* is defined as the following two parts: (1) Dialog-Level Doc Belief $p^t \in \mathbb{R}^M$ represents the probability of each document being selected. (2) Dialog-Level Attr Belief $\pi^t \in \mathbb{R}^L$ is the probability of an attribute being unknown. The lower the value, the higher the attribute belief.

When a document is excluded, it will rarely be selected. But the attribute belief is accumulated for each one, indicating whether the attribute will be asked. This two belief distributions are updated as follows.

$$p^t = \text{norm}(p^{t-1} \circ \hat{p}^t) \tag{14}$$

$$\pi^t = \min(\pi^{t-1} + \hat{\pi}^t, 1) \tag{15}$$

where norm is the L1-Normalization method. The initial value p^0 is a uniform distribution, and π^0 is initialized to zero vector. At the beginning of each dialogue, agent directly use p^0 and π^0 into the PM module.

DaPO for PM

We propose ***D**ocument-**a**ware **P**olicy **O**ptimizing (DaPO)* for PM to minimize the number of dialogue turns and guess the true target.

In order to describe the degree of discreteness of the data, we introduce a similar variance measure to calculate the

[1] Doc is the abbreviation of document and Attr is attribute.

differentiated representation $Q_i^{diff} \in \mathbb{R}^{L\times 4d}$ for each document Q_i. It's used to describe the degree of attribute discreteness.

$$\overline{Q} = \frac{1}{N}\sum_{i=1}^{N} Q_i \tag{16}$$

$$Q_i^{diff} = (Q_i - \overline{Q})^2 \tag{17}$$

where N is the size of whole dataset.

Note that the Dialog-Level Doc Belief p^t is the confidence of each document. Therefore, a weighted sum is used on the differentiated representation $Q^{diff} \in \mathbb{R}^{M\times(L\times 4d)}$ (concatenation of each candidate Q_i^{diff} etc.) to obtain $v^t \in \mathbb{R}^{L\times 4d}$, which is used to describe attribute uncertainty $\gamma^t \in \mathbb{R}^L$ over all document candidates.

$$v^t = (Q^{diff})^T p^t \tag{18}$$

$$\gamma^t = v^t W^{diff} \tag{19}$$

where $W^{diff} \in \mathbb{R}^{4d\times 1}$ is a trainable parameter.

Normally, the agent can directly ask the attribute with the largest uncertainty γ^t, expected to successfully end the dialogue in a minimum number of turns. However, some highly uncertain attributes may have low belief, therefore, the "*ask*" action of time step $t+1$ should be calculated as:

$$a^{t+1} = \mathrm{softmax}(\gamma^t(1-\pi^t)) \tag{20}$$

It is also interesting to note that a^{t+1} is equivalent to the probability of each attribute being asked at timestep $t+1$.

Rule for NLG

In NLG module, the agent generates natural language response respectively for "*ask*" and "*guess*" action. For "*ask*" action, it produces a response of asking the most probable attribute greedily according to a^{t+1}. For "*guess*" action, the agent guesses the movie corresponding with the most probable document greedily according to p^t. We use predefined natural language templates to converse with the user. The termination includes two cases.

1. Positive termination: when the maximum value of Dialog-Level Doc Belief p^t exceeds a certain threshold K.
2. Passive termination: when the set maximum number of dialogue turns is reached.

Experiments

Experimental Setting

We divide GuessMovie into two disjoint parts. The 70% part is used for pre-training document representation and NLU module, and the remaining 30% is used for training MD3 with 50k simulations using REINFORCE (Williams 1992) algorithm. The discount rate is 0.9. After training, we run a further 5k simulations to test the performance.

We construct a user simulator (Schatzmann et al. 2007; Li et al. 2016) with handcrafted rules because the user only need to answer the agent's questions passively. It randomly selects the target at the beginning. During the dialogue, the relevant value is indexed from the current structured KB and filled into a natural language template to response. Otherwise, if the user doesn't have knowledge, the answer is unknown. Specifically, in order to increase the dialogue diversity, we randomly mask some values for the 6 attributes with proportion of 0.1 at the beginning of each dialogue. However, all documents remain unchanged in the agent.

The reward function is similar to Dhingra et al. (2017). If the rank of target document is in the top $R = 3$ results, the reward is $max(0, 2(1-(r-1)/R))$, where r is the actual rank of the target. Otherwise, if the dialogue failes, the reward is -1. In addition, a reward of -0.1 will be given in each turns, making the dialogue tend to be finshed in a smaller number of turns.

We use Adam(Kingma and Ba 2014) optimizer with learning rate 0.001 and GloVe (Pennington, Socher, and Manning 2014) word embeddings. The number of candidate documents for document representation and dialogue is 32 by default. The maximum number of turns is 5. The probability threshold K for whether performing a Guess action is 0.5.

Baselines

The following introduces several different modules to compare with our method, including the machine reading comprehension (MRC) for NLU, random and fixed policy for PM.

NLU Here we introduce two methods.

- `MRC`: directly use the attributes and values for a document, which is extracted by BERT(Devlin et al. 2018) with no answer supported. The part of GuessMovie dataset used for documents pre-training are modified into a standard extractive MRC dataset, and divided into training, development and test parts. The testing EM score is 82.99 and F1 score is 87.44. After training, we extract the structured MRC-KB from the other part of GuessMovie dataset for dialogue. The similarity of each movie and the current user response is measured by calculating the overlap ratio, and normalized as the Turn-Level Doc Belief $\hat{p}^t$. The Turn-Level Attr Belief $\hat{\pi}^t$ can be obtained directly.
- `DaLU`: the method proposed in this paper.

PM Here we introduce three methods.

- `Rand`: randomly selecting an attribute to ask.
- `Fixed`: asking attributes in a fixed order.
- `DaPO`: the method proposed in this paper.

Human Performance Human performance is the ceiling of this task by assuming human can understand the candidate documents accurately, and calculate the truth document distribution. An attribute with the most discriminating information can be generated. In this scenario, we directly adopt structured KB and attribute-value pairs instead of natural language, in order to simulating accurate interaction.

- `Human-NLU`: we use a handcrafted module to directly match the current turn with each structured KB. If they match, the selection probability in Turn-Level Doc Belief

NLU	PM	32					64					128				
		S1	**S3**	**M**	**T**	**R**	**S1**	**S3**	**M**	**T**	**R**	**S1**	**S3**	**M**	**T**	**R**
MRC	Rand	.51	.72	.63	5.00	0.87	.40	.62	.56	5.00	0.57	.28	.53	.42	5.00	0.27
MRC	Fixed	.67	.89	.78	5.00	1.29	.57	.79	.69	5.00	1.06	.49	.67	.61	5.00	0.76
MRC	DaPO	.79	.94	.87	5.00	1.49	.71	.93	.82	5.00	1.41	.52	.86	.69	5.00	1.23
DaLU	Rand	.56	.81	.70	4.34	1.21	.47	.71	.62	4.64	0.91	.38	.61	.53	4.84	0.57
DaLU	Fixed	.71	.93	.83	3.78	1.57	.65	.87	.77	4.28	1.37	.56	.80	.70	4.66	1.14
DaLU	**DaPO**	**.83**	**.97**	**.90**	**3.42**	**1.73**	**.74**	**.93**	**.84**	**3.86**	**1.56**	**.67**	**.88**	**.78**	**4.29**	**1.38**
DaLU	w/o AU	.65	.91	.79	3.88	1.51	.62	.86	.75	4.42	1.33	.55	.81	.70	4.76	1.14
DaLU	w/o AB	.47	.82	.65	4.64	1.14	.45	.74	.61	4.48	0.98	.36	.63	.52	4.70	0.66
Human	Human	**.99**	**.99**	**.99**	**3.28**	**1.87**	**.98**	**.99**	**.98**	**3.52**	**1.83**	**.96**	**.99**	**.97**	**3.80**	**1.79**

Table 2: Dialogue test results with various combinations of NLU and PM on GuessMovie dataset. *S1* denotes the top-1 dialogue success rate. *S3* denotes the top-3 dialogue success rate. *M* denotes the target document ranking metric MRR. *T* denotes the average number of dialogue turns. *R* denotes the average rewards. *AU* denotes attribute uncertainty. *AB* denotes attribute belief.

$\hat{p}^t$ is set to 1, otherwise the probability is set to 0. It's the same for Turn-Level Attr Belief $\hat{\pi}^t$.

- `Human-PM`: the Dialog-Level Doc Belief p is thresholded to obtain a filtered documents. Based on the structured KB, the entropy of each attribute is calculated, and normalized as distribution γ^t. After that, it is still fused with the Dialog-Level Attr Belief π^t.

Results

We simulated 5k dialogues randomly for testing the performance. The complete results are shown in the Table 2. Various combinations of NLU and PM module are selected, and different size of candidate documents (*32, 64, and 128*) are setted. We calculated the dialogue success rate of the target in the candidates top-1 or top-3. In addition, similar to the retrieval task, the Mean Reciprocal Rank (MRR) is also calculated based on the position of the target in the ranking result. At the same time, the average number of dialogue turns and average rewards are also obtained.

As shown in the Table 2, the DaLU-NLU module has a higher dialogue success rate and smaller number of turns compared to the MRC-NLU module with the same PM, because it's difficult for the MRC model to accurately extract the attributes and there may be implied attributes value in the document. Therefore more interaction is needed to improve the select probability of the target.

In the combinations of DaLU-NLU, DaPO-PM is significantly superior to the random or fixed policy which is impossible to generate appropriate actions for specfic candidates, while DaPO-PM can ask attribute with higher uncertainty and belief.

Although MD3 has significant advantages compared with others and is very close to human's performance, it's performance reduced quickly when increasing the size of candidate documents to 64 or 128. Also, there is still a lot of room for improvement on top-1 dialogue success rate and larger size of candidate documents.

Ablation Study

Document Representation We use an Attribute-aware HAN (AaHAN) encoder for document representation, which accurately capture the key information for different attributes. By removing the attribute-aware mechanism, we only use the HAN encoder, which means each attribute share the same parameters. As shown in Table 3, AaHAN has a significant improvements which demonstrates the importance of sufficient document knowledge representation for such dialogue.

Encoder	**S1**	**S3**	**M**	**T**	**R**
AaHAN	**0.83**	**0.97**	**0.90**	**3.42**	**1.73**
HAN	0.33	0.63	0.52	4.89	0.67

Table 3: Dialogue test results for different encoder on document representation with candidate size of 32.

Dialog Policy To demonstrate the necessary of attribute uncertainty γ^t and attribute belief π^t for dialog policy, we introduce two ablation tests on PM.

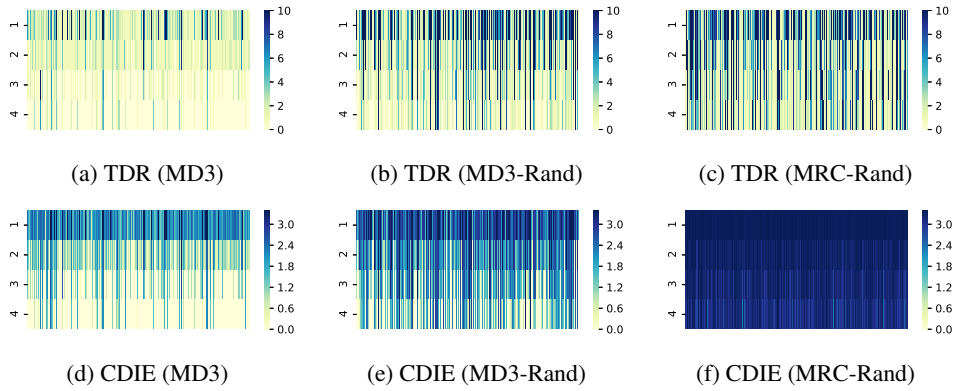
- `w/o AU`: DaPO without attribute uncertainty γ^t.
- `w/o AB`: DaPO without attribute belief π^t.

As shown in Table 2, both performance is significantly degraded over several metrics and different size of candidates. To make an accurately guess in shortest turns, we should not only consider the higher attribute uncertainty, but also the attribute belief.

Guess Threshold By adjusting the threshold K which decides whether to make a guess, different policies can be obtained, as shown in the Table 4. When the threshold K is large, it tends to make accurate guess and is not limited to the shortest dialogue turns. Otherwise, the top-1 document maybe not accurate, but the dialogue turns is shorter. This is a contradictory problem. We finally select threshold value of 0.5 to obtain a sub-optimal policy.

Analysis

At the end of each dialogue, a sorting results can be obtained according to the Dialog-Level Doc Belief p^t. For all 5k simulated dialogs, we visualize the dynamic change of the target document rank (TDR), as shown in 3-a. we can see that from top to bottom, as dialogue goes on, the color of the block



(a) TDR (MD3) (b) TDR (MD3-Rand) (c) TDR (MRC-Rand)

(d) CDIE (MD3) (e) CDIE (MD3-Rand) (f) CDIE (MRC-Rand)

Figure 3: Visualization of target document rank (TDR) and candidate documents information entropy (CDIE) dynamic changes. The ordinate represents the end of each dialogue turns. MD3-Rand and MRC-Rand denote DaLU or MRC with Rand policy.

Turn	Dialog(MD3)	R/P	Dialog(MD3-Rand)	R/P	Dialog(MRC-Rand)	R/P
0	-	23/ 0.03	-	23/ 0.03	-	23/ 0.03
1	When is it released? It's a 1955 film.	2/ 0.17	What's the language? The language is english.	10/ 0.11	What's the release year? It's a 1955 film.	3/ 0.07
2	What's the director? Charles Vidor is the director.	1/ 0.60	What's the director? Charles Vidor is the director.	3/ 0.23	What's the genre? It's a drama.	18/ 0.03
3	I guess the movie is love_me_or_leave_me. Your guess is True.	-	Does it has any other language? I don't know the answer.	3/ 0.23	What's the language? The language is english.	17/ 0.03
4	-	-	When is it released? It's a 1955 film.	1/ 0.72	Does it has any other language? I don't know the answer.	17/ 0.03
5	-	-	I guess the movie is love_me_or_leave_me. Your guess is True.	-	I guess the movie is beethoven's_3rd. Your guess is False.	-

Figure 4: Examples of different dialogue models with the same candidate documents. R and P means target document rank and probability after each turn. The darker the color, the higher the rank, the greater the selection probability.

K	S1	S3	M	T	R
0.5	0.8298	0.9708	0.901	**3.42**	**1.73**
0.7	0.8756	0.9726	0.926	3.88	1.71
0.9	**0.8774**	**0.9728**	**0.927**	4.44	1.66

Table 4: Dialogue test results for different document guess thresholds with candidate size of 32.

is gradually lighter, which means TDR is gradually higher. In addition, the candidate documents information entropy (CDIE) calculated by the Dialog-Level Doc Belief p^t can also be visualized, indicating uncertainty change of the selected document, as shown in 3-b. From top to bottom, as dialogue goes on, the CDIE gradually decreases, which means the uncertainty of the guessed document become smaller. It illustrates that our model is interpretable.

In the comparison of several combinations, we can find that MD3 has significant advantages regardless of TDR and CDIE.

Case Study

Figure 4 shows three dialogue samples between different models with the same candidates, and dynamic changes of the target rank and probability.

At the beginning of the dialogue, each document has the same probability to be guessed as the target. In the MD3 sample, the "*release year*" and "*directed by*" attributes are asked, so that the rank of the target document quickly rises to the first place, and the probability value increases to 0.6. As for the MRC-Rand sample, it doesn't ask the director, which has the greatest difference for candidates. And the probability of selecting the target is always low. This is due to the inaccuracy of the MRC model for attributes extraction and implied attribute information. It can be seen that MD3 is better than the other methods and more robust.

Conclusions

In this paper, we introduced a new multiple document-driven dialogue task, and proposed a public benchmark dataset—GuessMovie. Besides, we investigated a multiple document-driven dialogue model which can converse with the user and achieve the dialogue goal conditioned on both document engagement and user feedback. Additionally, although our model has significant advantages over several strong baselines. We hypothesize that there are several pre-defined attributes for documents and agent can only ask such questions. How to extend the dialogue to more scenarios with less restriction is a further research.

Acknowledgments

We thank the anonymous reviewers for their insightful comments. The research is supported in part by the National Key Research and Development Program of China under the grant of 2020YFF0305302.

References

Burgener, R. 2006. Artificial neural network guessing method and game. US Patent App. 11/102,105.

Das, A.; Kottur, S.; Moura, J. M.; Lee, S.; and Batra, D. 2017. Learning cooperative visual dialog agents with deep reinforcement learning. In *Proceedings of the IEEE international conference on computer vision*, 2951–2960.

De Vries, H.; Strub, F.; Chandar, S.; Pietquin, O.; Larochelle, H.; and Courville, A. 2017. Guesswhat?! visual object discovery through multi-modal dialogue. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 5503–5512.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* .

Dhingra, B.; Li, L.; Li, X.; Gao, J.; Chen, Y.-N.; Ahmad, F.; and Deng, L. 2017. Towards End-to-End Reinforcement Learning of Dialogue Agents for Information Access. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 484–495.

Dinan, E.; Roller, S.; Shuster, K.; Fan, A.; Auli, M.; and Weston, J. 2018. Wizard of wikipedia: Knowledge-powered conversational agents. *arXiv preprint arXiv:1811.01241* .

Ghazvininejad, M.; Brockett, C.; Chang, M.-W.; Dolan, B.; Gao, J.; Yih, W.-t.; and Galley, M. 2018. A knowledge-grounded neural conversation model. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

Hadsell, R.; Chopra, S.; and LeCun, Y. 2006. Dimensionality reduction by learning an invariant mapping. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 2, 1735–1742. IEEE.

Hu, H.; Wu, X.; Luo, B.; Tao, C.; Xu, C.; Wu, W.; and Chen, Z. 2018. Playing 20 Question Game with Policy-Based Reinforcement Learning. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 3233–3242.

Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* .

Li, X.; Lipton, Z. C.; Dhingra, B.; Li, L.; Gao, J.; and Chen, Y.-N. 2016. A user simulator for task-completion dialogues. *arXiv preprint arXiv:1612.05688* .

Li, Z.; Niu, C.; Meng, F.; Feng, Y.; Li, Q.; and Zhou, J. 2019. Incremental Transformer with Deliberation Decoder for Document Grounded Conversations. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 12–21.

Liu, H.; Yuan, C.; and Wang, X. 2020. Label-Wise Document Pre-training for Multi-label Text Classification. In *CCF International Conference on Natural Language Processing and Chinese Computing*, 641–653. Springer.

Madotto, A.; Wu, C.-S.; and Fung, P. 2018. Mem2Seq: Effectively Incorporating Knowledge Bases into End-to-End Task-Oriented Dialog Systems. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1468–1478.

Miller, A.; Fisch, A.; Dodge, J.; Karimi, A.-H.; Bordes, A.; and Weston, J. 2016. Key-Value Memory Networks for Directly Reading Documents. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 1400–1409.

Pang, W.; and Wang, X. 2020a. Guessing State Tracking for Visual Dialogue. In *ECCV*.

Pang, W.; and Wang, X. 2020b. Visual Dialogue State Tracking for Question Generation. In *AAAI*.

Parthasarathi, P.; and Pineau, J. 2018. Extending Neural Generative Conversational Model using External Knowledge Sources. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 690–695.

Pennington, J.; Socher, R.; and Manning, C. D. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 1532–1543.

Reddy, S.; Chen, D.; and Manning, C. D. 2019. CoQA: A Conversational Question Answering Challenge. *Transactions of the Association for Computational Linguistics* 7: 249–266.

Schatzmann, J.; Thomson, B.; Weilhammer, K.; Ye, H.; and Young, S. 2007. Agenda-based user simulation for bootstrapping a POMDP dialogue system. In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Companion Volume, Short Papers*, 149–152. Association for Computational Linguistics.

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

Wu, C.-S.; Socher, R.; and Xiong, C. 2019. Global-to-local Memory Pointer Networks for Task-Oriented Dialogue. In *Proceedings of the International Conference on Learning Representations (ICLR)*.

Yang, Z.; Yang, D.; Dyer, C.; He, X.; Smola, A.; and Hovy, E. 2016. Hierarchical attention networks for document classification. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies*, 1480–1489.

Zhao, T.; and Eskenazi, M. 2016. Towards End-to-End Learning for Dialog State Tracking and Management using Deep Reinforcement Learning. In *17th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 1.