Balanced Meta Learning and Diverse Sampling for Lifelong Task-Oriented Dialogue Systems

Qiancheng Xu[1], **Min Yang**[2*], **Ruifeng Xu**[3]
[1]Georgia Institute of Technology
[2]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences
[3]Harbin Institute of Technology (Shenzhen)
qxu309@gatech.edu, min.yang@siat.ac.cn, xuruifeng@hit.edu.cn

Abstract

In real-world scenarios, it is crucial to build a lifelong task-oriented dialogue system (TDS) that continually adapts to new knowledge without forgetting previously acquired experiences. Existing approaches mainly focus on mitigating the catastrophic forgetting in lifelong TDS. However, the transfer ability to generalize the accumulated old knowledge to new tasks is underexplored. In this paper, we propose a two-stage lifelong task-oriented dialogue generation method to mitigate catastrophic forgetting and encourage knowledge transfer simultaneously, inspired by the learning process. In the first stage, we learn task-specific masks which adaptively preserve the knowledge of each visited task so as to mitigate catastrophic forgetting. In this stage, we are expected to learn the task-specific knowledge which is tailored for each task. In the second stage, we bring the knowledge from the encountered tasks together and understand thoroughly. To this end, we devise a balanced meta learning strategy for both forward and backward knowledge transfer in the lifelong learning process. In particular, we perform meta-update with a meta-test set sampled from the current training data for forward knowledge transfer. In addition, we employ an uncertainty-based sampling strategy to select and store representative dialogue samples into episodic memory and perform meta-update with a meta-test set sampled from the memory for backward knowledge transfer. With extensive experiments on 29 tasks, we show that MetaLTDS outperforms the strong baselines in terms of both effectiveness and efficiency. For reproducibility, we submit our code at: https://github.com/travis-xu/MetaLTDS.

Introduction

Task-oriented dialog systems (TDSs) allow users to accomplish complex tasks using natural language in an interactive manner, such as restaurant reservations, car navigation, and weather inquiry. Existing end-to-end TDSs (Peng et al. 2021; Hosseini-Asl et al. 2020) have achieved impressive performance on specific domains in a static environment, where the datasets from different domains remain unchanged and are always available. However, a TDS in the production environment often encounters continuous streams of domains or functionalities through time (Geng et al. 2021; Madotto et al. 2021). The TDSs with isolated learning (Lee 2017) would suffer from significant performance degradation on previously learned tasks after being trained on a new one, which is known as catastrophic forgetting (McCloskey and Cohen 1989; Goodfellow et al. 2014). In addition, existing TDSs struggle to be adapted to new domains or functionalities so as to support new services over time (Mazumder and Liu 2022). Therefore, it is crucial for a TDS to continually learn new knowledge without forgetting previously acquired functionalities when confronted with a sequence of different tasks, which is known as lifelong or continual learning (Biesialska, Biesialska, and Costa-jussà 2020; Qu et al. 2021).

One major challenge in lifelong learning is to balance the trade-off between stability and plasticity (i.e., the stability-plasticity dilemma) (Mermillod, Bugaiska, and Bonin 2013). On the one hand, although a TDS is expected to remember previously acquired knowledge and avoid catastrophic forgetting (i.e., high stability), focusing too much on stability may hinder it from quickly adapting to new tasks. On the other hand, when a TDS is drastically updated based on new knowledge (i.e., high plasticity), it may quickly forget previously-learned experiences and thus suffer from catastrophic forgetting of previous tasks.

In recent years, plenty of lifelong learning studies have been proposed, which can be divided into three categories: memory-based approaches (de Masson d'Autume et al. 2019), architecture-based approaches (Mallya and Lazebnik 2018), and regularization-based approaches (Kirkpatrick et al. 2017; Ahn et al. 2019). These methods primarily focus on dealing with the catastrophic forgetting problem in lifelong learning, which preserve the previously learned knowledge to hold long-term durability when learning new tasks over time. However, how to effectively reuse previously acquired knowledge when processing new tasks remains underexplored. One possible solution is to employ knowledge transfer (e.g., domain adaptation) techniques to adapt a model trained on source domains to target domains, exploiting the previously learned knowledge in future learning. The performance of domain adaptation techniques (Volpi, Larlus, and Rogez 2021) generally depends on the similarity between source and target data distributions. Nevertheless, in the lifelong learning scenario, we cannot decide whether the sequentially coming domains will be more or less similar

*Min Yang is corresponding author.

to each other. In addition, aligning the feature distributions of multi-source domains and the target domain data distribution can even worsen the performance of the TDS without feature distribution alignment, which is referred to as "negative transfer". Thus, it is non-trivial to employ the standard domain adaptation techniques for lifelong learning.

In this paper, we propose a novel two-stage lifelong learning method (called MetaLTDS) for task-oriented dialogue generation, which simultaneously mitigates the catastrophic forgetting problem and encourages knowledge transfer from old tasks to new tasks. In the first stage, we learn task-specific masks which adaptively preserve the knowledge of each visited task so as to overcome catastrophic forgetting. In the second stage, we devise a balanced meta learning strategy aiming at promoting both forward and backward knowledge transfer. Concretely, we first perform meta-update with a bunch of current task samples (denoted as the support set) for forward transfer. In addition, we devise an uncertainty-based sampling strategy to select and store representative dialogue samples in episodic memory, and update the model by explicitly optimizing the loss with the meta-test set sampled from the memory for backward transfer. We further add a balancing factor to the overall loss function for controlling the importance of old and new knowledge. Note that the updates are conducted on all the preserved (masked) parameters, without introducing extra model capacity. In this manner, the parameters preserved (masked) for visited tasks are forced to adapt to the new task (i.e., forward knowledge transfer), while new parameters would learn to generalize on the experiences grasped from preceding tasks (i.e., backward knowledge transfer).

Our main contributions can be summarized as follows:

- We introduce a novel lifelong task-oriented dialogue system with adaptive masking and balanced meta learning, which alleviates catastrophic forgetting while promoting both forward and backward knowledge transfer.
- We propose an uncertainty-based sampling method to select limited representative and diverse dialogue samples for the visited tasks and then store these samples in the episodic memory, which are then used as the meta-test for effective backward knowledge transfer.
- We conduct extensive experiments on three datasets under a disjoint-domain setup and a novel blurry-domain setup. Experimental results demonstrate that MetaLTDS outperforms baselines with high parameter efficiency.

Related Work

Task-oriented Dialog Systems TDSs allow users to seek information and complete complex tasks using natural language in an interactive manner. Recently, increasing attention has been gained to build end-to-end TDSs by directly mapping dialogue history to target responses without requiring hand-crafted state labels. For example, Hosseini-Asl et al. (2020) proposed SimpleTOD based on the GPT-2 (Radford et al. 2019) model which generated target sequences for accomplishing each sub-task. Peng et al. (2021) devised SOLOIST which was pre-trained on external dialogue corpora to improve the downstream performance. GALAXY (He et al. 2022b) and SPACE (He et al. 2022a) introduced semi-supervised dialogue model pre-training, which achieved impressive results on several datasets.

Lifelong Task-oriented Dialogue Systems When a stream of domains or functionalities come sequentially, it is essential to develop a lifelong TDS. Recently, several lifelong learning methods have been proposed to overcome catastrophic forgetting. Geng et al. (2021) proposed a continual learning approach for TDSs with iterative network pruning, expanding and masking. ToDCL (Madotto et al. 2021) is the most relevant work to our method, which provides a strong benchmark for end-to-end TDSs. However, ToDCL neglects the knowledge transfer between tasks since it blindly adds a separate and independent adapter (Houlsby et al. 2019) to the pre-trained backbone model for each new task. Different from ToDCL, our method enables both forward and backward knowledge transfer via meta learning, and further reduces the parameter redundancy.

Lifelong and Continual Learning Lifelong learning, also called continual learning, aims to acquire knowledge from a sequence of tasks, without forgetting previously learned tasks (Biesialska, Biesialska, and Costa-jussà 2020; Qu et al. 2021). Generally, previous lifelong learning approaches mainly focused on mitigating catastrophic forgetting, which can be divided into three categories: (i) regularization-based approaches that add additional constraints on tasks' objectives to consolidate previously learned knowledge (Kirkpatrick et al. 2017; Li and Hoiem 2018; Ahn et al. 2019), (ii) rehearsal-based approaches that store some training samples from old tasks and replay them in learning new tasks (de Masson d'Autume et al. 2019; Chaudhry et al. 2019a), and (iii) architecture-based approaches that allocate specific parameters to different tasks and keep old parameters fixed when learning a new task (Mallya and Lazebnik 2018; Hung et al. 2019; Yan, Xie, and He 2021). In this paper, we simultaneously mitigate catastrophic forgetting and encourage knowledge transfer by taking advantage of both architecture-based and rehearsal-based approaches.

Meta learning The goal of meta learning is to equip the model with the ability to rapid adaptation to unseen tasks. MER (Riemer et al. 2019) conducted meta-learning based on gradients to maximize knowledge transfer and minimize task interference. iTAML (Rajasegaran et al. 2020) learned a task-agnostic model, which first identified and then adapted to the target task through meta-update. Meta-DR (Volpi, Larlus, and Rogez 2021) further proposed a meta-learning strategy for continual domain adaptation, which devised a two-fold regularizer to restrain domain shift and ease domain adaptation. In this work, we leverage a balanced meta learning method to boost knowledge transfer, where the current model could not only obtain knowledge from previous tasks but also "learns to learn" future tasks as well.

Problem Setting

Task-oriented Dialogue Generation In this paper, a TDS is formulated as an end-to-end generation task, where a sequence-to-sequence (Seq2Seq) model is employed to

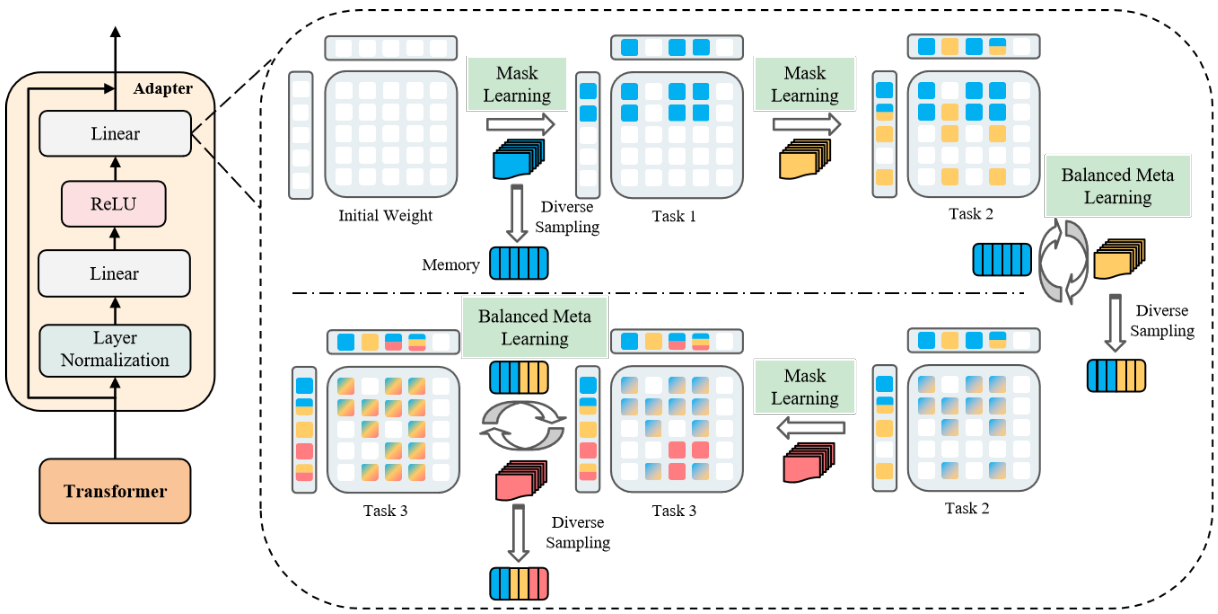

Figure 1: The overview of MetaLTDS. The neurons in blue, yellow and red are tailored for tasks 1, 2 and 3, respectively.

generate the target sequence (system response) $Y = \{y_1, ..., y_m\}$ word by word given the dialogue history $X = \{x_0, ..., x_n\}$ as input.

Lifelong Task-oriented Dialogue Generation The goal of a lifelong TDS is to build a system that continually learns over time by grasping new knowledge and retaining previously learned experiences. Formally, a lifelong TDS is expected to handle a sequence of T tasks $\{\mathcal{T}_1, ..., \mathcal{T}_T\}$, where each task $\mathcal{T}_t$ contains a bunch of conversations spanning over a set of dialogue domains D_t. We denote the multi-turn dialogues for the t-th task as $S_t = \{(X_i, Y_i)_{i=1}^N\}$, where N is the number of dialogues.

We introduce two setups for model evaluation, i.e., a disjoint-domain setup and a blurry-domain setup, according to whether the current dialogue domain D_t intersects with previously encountered dialogue domains $\{D_1, ..., D_{t-1}\}$:

$$(D_1 \cup ... \cup D_{t-1}) \bigcap D_t = \varnothing, \quad \text{(Disjoint)}$$
$$(D_1 \cup ... \cup D_{t-1}) \bigcap D_t \neq \varnothing, \quad \text{(Blurry)}$$

The disjoint-domain setup is a common practice for evaluating lifelong TDSs. Since different tasks have no overlapped domains, the shared knowledge between tasks is limited and the catastrophic forgetting problem would be exaggerated (Ke et al. 2021). To further demonstrate the knowledge transferability of our lifelong TDS, we draw inspiration from (Bang et al. 2021; woo Koh et al. 2021) and propose a blurry-domain setup. In this blurry-domain setup, the task boundaries are faint where a newly coming task will contain at least one shared domain with previous tasks. In general, the shared knowledge can be beneficial for knowledge transfer from old tasks to the new task. In addition, the blurry-domain setup could be more realistic than the disjoint-domain setup since the domains covered in the old tasks may still appear in the subsequent tasks in real-world applications.

Our Methodology

Following (Madotto et al. 2021), we employ GPT-2 as our backbone model, along with residual adapters (Houlsby et al. 2019) for task parameterization. The residual adapter is trainable, and is inserted into each pre-trained transformer layer (Vaswani et al. 2017). As illustrated in Figure 1, an adapter module consists of a layer normalization (Ba, Kiros, and Hinton 2016), followed by two feature projection layers with a residual connection (He et al. 2016). During the training process of each task, only the adapters are updated while all the other GPT-2 parameters are frozen. This is a useful strategy for lifelong learning since fine-tuning the whole network may cause catastrophic forgetting of old tasks, similar to ToDCL (Madotto et al. 2021). However, ToDCL (Madotto et al. 2021) incorporates a new task-specific adapter into all the transformer layers for each newly coming task, which may hinder the knowledge transfer. In addition, model expansion is prohibited when being deployed to real-time applications with a large number of sequential tasks and resource-constrained devices. On the contrary, we keep learning multiple tasks with the same adapter module, enabling knowledge transfer via parameter reuse and meta learning. A new adapter is added only when there are not enough parameters to learn subsequent tasks, i.e., the size of the used parameters is close to the size of the adapter.

In this paper, we propose a two-stage lifelong learning method for task-oriented dialogue systems, inspired by the learning process of humans. As illustrated in Figure 1, in the first stage, we learn task-specific masks which adaptively preserve the knowledge of each visited task so as to mitigate catastrophic forgetting. In this stage, we can learn the task-specific knowledge which is tailored for each task. In the second stage, we integrate the knowledge from the encountered tasks by devising a balanced meta learning strategy, aiming at enabling both forward and backward knowledge transfer. Next, we will introduce these two stages in detail.

Stage 1: Adaptive Mask Learning

In the first stage, we learn task-specific masks which adaptively preserve the knowledge of each visited task so as to mitigate catastrophic forgetting. In particular, the task-specific masks are applied to the output of the two feature projection layers of the adapters, indicating which neurons should be preserved for the individual tasks so as to overcome catastrophic forgetting and knowledge interference (Serra et al. 2018). When learning the masks for subsequent tasks, we keep the preserved neurons fixed by preventing the corresponding parameters from being updated during back-propagation. We allow overlapping masks between tasks to encourage the reuse of previously selected neurons. The overlapping masks indicate that these features can be transferred and shared between tasks. In this manner, we could reduce the capacity usage and reserve the network space for future tasks. If the percentage of masked neurons has reached the adapter capacity (predefined by an upper limit $p\%$), then we add a new adapter for subsequent tasks.

Illustration The mask learning process is illustrated in Figure 1. The feature masks of two feature projection layers are represented by the tensors on the left and top of the weight matrix respectively. At the beginning of the learning process, all neurons in the matrices are initialized at random, which are colored in white and have not been used by any task yet. After learning the first task, the blue neurons in the masks indicate the task-specific features tailored for the first task, which are then preserved for the first task. After learning the second task, the neurons with two colors suggest that there exist overlapping masks (knowledge) between the first and second tasks. The weight parameters that are masked could be reused by the corresponding tasks in learning subsequent tasks. Note that these masked neurons are designated for the next learning stage via balanced meta learning, without introducing extra parameters.

Stage 2: Balanced Meta Learning

In the second stage, we aim to bring the knowledge learned from the visited tasks together and understand thoroughly, which could be beneficial to future learning as well. To this end, we borrow the idea in meta learning literature (Volpi, Larlus, and Rogez 2021) for both forward and backward knowledge transfer in the lifelong learning process. Different from typical model training that updates the model parameters directly, meta learning involves a meta-update step prior to the final update. The two steps are conducted on a meta-support set S_{sup} and a meta-test set S_{test}, respectively. For meta-update, we first sample a mini-batch from S_{sup} and calculate its loss associated with the initial model θ. Then the meta-updated model $\hat{\theta}$ could be derived by $\hat{\theta} = \theta - \alpha \nabla_\theta \mathcal{L}_{sup}(\theta)$, where $\mathcal{L}_{sup}(\theta)$ is the loss for the mini-batch from S_{sup} and α is the learning rate for meta-update. $\hat{\theta}$ simulates what the model would turn in future learning. Then, we perform the final update on θ as $\theta = \theta - \eta \nabla_\theta \mathcal{L}_{test}(\hat{\theta})$, where $\mathcal{L}_{test}(\hat{\theta})$ denotes the loss calculated on mini-batches from S_{test} with respect to $\hat{\theta}$. Here, η is the learning rate for the final update. In this manner, we can provide a better direction for updating the current model by optimizing the performance of the meta model (simulated by $\hat{\theta}$).

In this paper, we introduce a balanced meta learning method to update the model for knowledge transfer. To better exploit the important information from visited tasks, we devise an uncertainty-based sampling strategy (described in Section 4.3) to select and store representative dialogue samples for each visited task in episodic memory, which can be used for knowledge transfer. Formally, after learning $t-1$ tasks, we have a memory buffer M_{t-1} that stores a limited number of representative samples for each visited task. The training samples S_t from the current task t are exploited as meta-support set S_{sup} to perform meta-update throughout the training stage, while the meta-test set S_{test} is sampled from the memory set M_{t-1} together with the current training set S_t.

Our meta learning strategy is performed as follows. First, we unfreeze all previously preserved weights and make all the parameters (i.e., the parameters for both old tasks and the current task) learnable. Then, we sample H mini-batches randomly from the meta-support set S_{sup}. For each mini-batch, we conduct a single meta-update step on the current model θ. Then we can get H different meta-updated models, which could jointly find the optimal direction for updating θ. For final update, we first sample two mini-batches $(\mathrm{x_n}, \mathrm{y_n})$ and $(\mathrm{x_o}, \mathrm{y_o})$ from S_{test}. Here, $(\mathrm{x_n}, \mathrm{y_n})$ and $(\mathrm{x_o}, \mathrm{y_o})$ denote samples from current dataset S_t and memory M_{t-1}, respectively. We perform final update with respect to θ via two loss functions computed on meta-updated models associated with $(\mathrm{x_n}, \mathrm{y_n})$ and $(\mathrm{x_o}, \mathrm{y_o})$. By enforcing the loss associated with $(\mathrm{x_n}, \mathrm{y_n})$ to be small, previously preserved old neurons (fixed by masks) could be adapted to new knowledge. On the other hand, minimizing the loss associated with $(\mathrm{x_o}, \mathrm{y_o})$ can encourage current neurons to accumulate old knowledge, promoting backward transfer. Following (Volpi, Larlus, and Rogez 2021), we only perform a single meta-optimization step. After e iterations of the inner-outer loop procedure of meta learning, we obtain the final model θ^t for the current task t. Note that this learning process is only conducted on all the masked neurons from task 1 to t (see Figure 1), without causing extra capacity.

As more tasks arrive, it would be more challenging to perform generalization on past knowledge since the scale of old neurons keeps expanding. Taking this factor into account, we add a balancing factor β on the loss functions to regulate their impact in the meta-optimization. Then, the loss associated with $(\mathrm{x_o}, \mathrm{y_o})$ would be multiplied by a larger factor as the number of visited tasks increases. In this manner, the model could attach more importance to model generalization.

The whole learning procedure is detailed in Algorithm 1.

Algorithm 1: Balanced Meta Training

Input: current task id t, current training set S_t, memory buffer M_{t-1}, initial weights θ^{init}, η (learning rate), α (meta learning rate), β (balancing factor)
Output: weights θ^t
Initialize: $\theta \leftarrow \theta^{init}, \beta \leftarrow 1/t$

for e iterations **do**
  for $h \in [1, H]$ **do**
    Sample $(\hat{\mathrm{x}}, \hat{\mathrm{y}})_h$ uniformly from S_t
    $\hat{\theta}^h \leftarrow \theta - \alpha \nabla_\theta (\mathcal{L}_\mathcal{T}(\hat{\mathrm{x}}_h, \hat{\mathrm{y}}_h; \theta))$
  end for
  Sample $(\mathrm{x_o}, \mathrm{y_o})$ uniformly from M_{t-1}
  Sample $(\mathrm{x_n}, \mathrm{y_n})$ uniformly from S_t
  $\theta \leftarrow \theta - \eta \nabla_\theta \sum_h \big(\beta \underbrace{\mathcal{L}_\mathcal{T}(\mathrm{x_n}, \mathrm{y_n}; \hat{\theta}^h)}_{\text{Forward transfer}} + (1-\beta) \underbrace{\mathcal{L}_\mathcal{T}(\mathrm{x_o}, \mathrm{y_o}; \hat{\theta}^h)}_{\text{Backward transfer}} \big)$
end for
$\theta^t \leftarrow \theta$

Uncertainty-Based Diverse Sampling

For each encountered task, we select a subset of representative as well as diverse samples and store them in the memory buffer. In particular, the memory buffer we maintain has a fixed capacity of K samples, which does not grow as new tasks come sequentially. Then, the number of samples for the t-th task is given by $K_t = [K/t]$, where t is the number of visited tasks. Since the exemplars retained for each task would become fewer as more tasks arrive, it is necessary to maintain sufficient information of old tasks by improving the quality of stored samples in the memory buffer.

In this paper, we propose an uncertainty-based diverse sampling strategy (Bang et al. 2021) to select samples that are representative and diverse in the feature space. The strategy is based on the assumption that the chosen samples with high quality should meet two standards. On the one hand, the samples that are close to the center of the data distribution should be selected since they are representative of the task. On the other hand, the samples residing in a region that is close to the task boundary should be qualified as well because they are distinguishable from samples of other tasks. To satisfy both requirements, we explore uncertainty to approximate the locations of each sample in the data distribution. The uncertainty of a sample is measured according to its robustness against perturbations. Since the samples located near the task boundary are prone to be perturbed (Bang et al. 2021), the model outputs of them would vary drastically and result in higher uncertainty.

The perturbation and transformation on samples are implemented by data augmentation methods. Conventional augmentation methods that are widely adopted in NLP often refer to word substitution or back-translation. In contrast, we borrow a simple but effective idea from contrastive learning (Gao, Yao, and Chen 2021), which leverages the standard dropout technique for data augmentation. Specifically, we

Algorithm 2: Uncertainty-Based Diverse Sampling

Input: memory size K, current task id t, current training set S_t, memory set M_{t-1}, dialogue domains $D_1, ..., D_t$
Output: updated memory M_t

$M_t = \{\}$
$K_t = floor(K/t)$
for $i = 1, 2, ..., t$ **do**
  $M_i = \{(X, Y)_d | (X, Y) \in M_{t-1} \cup S_t, d \in D_i\}$
  Sort M_i by $u(X, Y)$ computed by Eq. 2
  for $q = 1, 2, ..., K_t$ **do**
    $p = q * |M_i| / K_t$
    M_t **+=** $M_i[p]$
  end for
end for

quantify the output of each perturbed pair (X, Y) as follows:

$$p(X,Y) = \sqrt[m]{\prod_{i=1}^{m} \frac{1}{p_\theta(y_i | x_0, ..., x_n, y_0, ..., y_{i-1})}} \tag{1}$$

where $p_\theta(y_i)$ denotes the prediction probability of a target output word y_i and $p(X, Y)$ is the average predictive distribution entropy across the whole target sequence.

Since dropout masks are randomly generated at each time, the output of a sample would be perturbed differently. Then, we compute the uncertainty of the sample as the average output after applying the random dropout A times:

$$u(X,Y) = 1 - \frac{1}{A} \sum_{i=1}^{A} p(X,Y) \tag{2}$$

A lower $u(X, Y)$ value indicates that (X, Y) contains more representative information for the corresponding task. After sorting all the dialogues by uncertainty, we select the top K_t samples for task t and store them in the memory buffer. Algorithm 2 summarizes the diverse sampling method.

Experimental Setup

Datasets

We evaluate our MetaLTDS approach on a total of 29 tasks from three benchmark TDS datasets under two setups. For the disjoint-domain setup, we conduct experiments on Task-Master 2019 (TM19) and Task-Master 2020 (TM20) datasets (Byrne et al. 2019) with 13 tasks. We randomly sample 2000 examples to form the training set of each task, and the ratio of train/val/test samples is kept at 8:1:1. For the blurry-domain setup, we conduct experiments on 16 multi-domain TDS tasks drawn from the Schema Guided Dialogue (SGD) dataset (Rastogi et al. 2020), which is randomly split into train/val/test samples for each task at the ratio of 7:1:2. All the experiments are conducted under the end-to-end setting (Madotto et al. 2021). The details of data statistics are provided in the Appendix (Table 4).

	Disjoint-domain Setup				Blurry-domain Setup				
Method	Intent ACC ↑	JGA ↑	BLEU ↑	EER ↓	Intent ACC ↑	JGA ↑	BLEU ↑	EER ↓	#Parameters ↓
Finetune	11.11	30.50	7.88	47.25	28.23	3.52	15.87	25.37	124.41M
EWC	10.60	31.50	8.91	40.99	28.26	3.53	15.79	25.26	124.41M
AGEM	23.60	32.91	8.18	46.53	34.95	3.82	16.83	23.18	124.41M
LAMOL	77.85	22.71	6.17	43.11	40.26	3.15	10.87	**17.67**	124.41M
Replay	77.34	46.90	12.68	37.58	55.19	5.53	19.04	18.91	124.41M
ToDCL	95.09	49.46	16.00	42.91	88.97	10.29	18.32	20.87	24.34M/29.95M
MetaLTDS	**95.40**	**52.73**	**18.15**	**37.42**	**89.24**	**12.27**	**19.93**	17.84	**16.72M**

Table 1: Average performance in terms of intent accuracy (Intent ACC), Joint-Goal-Accuracy (JGA), BLUE, and Slot-Error-Rate (EER) on the disjoint-domain setup and the blurry-domain setup. Note that the average parameters of ToDCL are not of the same size under the two setups due to different task numbers.

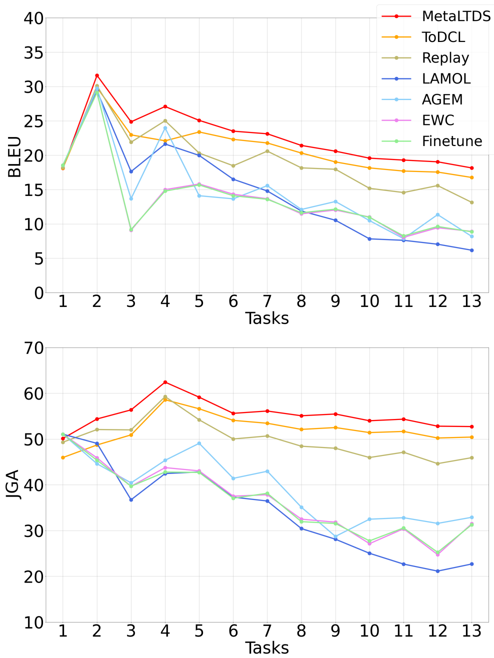

Figure 2: The average results in terms of BLEU and JGA over the 1-st to k-th tasks after learning the k-th task ($t \in [1, 13]$).

Baselines

We compare MetaLTDS with the following strong baselines: (1) **Finetune** that fine-tunes the GPT-2 model on sequential tasks, (2) **EWC** (Kirkpatrick et al. 2017) that uses a regularization term to prevent important parameters from updating too much, (3) **AGEM** (Chaudhry et al. 2019b) that is an efficient variant of the gradient episodic memory method, (4) **LAMOL** (Sun, Ho, and yi Lee 2020) that trains the model with current data and samples from pseudo experience replay, (5) **Replay** that maintains a memory buffer randomly sampled from past tasks and uses them to jointly train the current model, (6) **ToDCL** (Madotto et al. 2021) that freezes the pre-trained GPT-2 model and inserts a residual adapter for each task independently. Implementation details of MetaLTDS and baseline methods can be found in the Appendix.

Evaluation Metrics

Following the previous work (Madotto et al. 2021), we define four metrics for model evaluation: (1) Intent ACC: the average accuracy between the predicted intents and the gold intents. (2) JGA: the average Joint Goal Accuracy (JGA) (Wu et al. 2019) between the predicted dialogue states and the target dialogue states. (3) BLEU (Papineni et al. 2002): the n-gram (i.e., $n = 4$) overlap rate between the generated responses and the ground-truth responses. (4) EER (Wen et al. 2015): the slot error rate reflecting the ratio of values not appearing in the response. To measure the continual learning performance, we calculate a commonly used average metric (Lopez-Paz and Ranzato 2017) as follows:

$$\text{Avg.} = \frac{1}{T}\sum_{t=1}^{T}\Omega_{T,t} \tag{3}$$

where Ω could be Intent ACC, JGA, BLEU or EER scores. $\Omega_{T,i}$ is test performance of task i after learning all T tasks.

Following (Lopez-Paz and Ranzato 2017), we further design two metrics to evaluate the effectiveness of the dialogue systems for backward knowledge transfer (FWT) and forward knowledge transfer (BWT) respectively:

$$\text{FWT} = \frac{1}{T-1}\sum_{i=2}^{T}\Omega_{i-1,i},\ \text{BWT} = \frac{1}{T-1}\sum_{i=1}^{T-1}\Omega_{T,i} - \Omega_{i,i} \tag{4}$$

Experimental Results

Overall Performance Comparison

We conduct experiments on 13 sequential tasks with the disjoint-domain setup and 16 sequential tasks with the blurry-domain setup. We run all methods with the same ordering during the training process. We report the average testing results after all the tasks are visited, where each model keeps learning a new task by leveraging the weights learned from the old tasks as initialization. Table 1 summarizes the overall performance of our MetaLTDS method and the compared baselines. We can observe that Finetune, EWC and AGEM perform much worse than other methods, which

	JGA		BLEU	
Method	FWT	BWT	FWT	BWT
Finetune	21.82	-31.31	4.74	-12.09
LAMOL	13.04	-29.45	3.13	-12.23
Replay	22.73	-14.65	5.40	-7.67
ToDCL	22.95	+0.00	4.82	+0.00
MetaLTDS	**22.98**	**+0.43**	**5.44**	**+0.86**

Table 2: Forward and backward knowledge transfer performance in terms of JGA and BLEU.

suffer from catastrophic forgetting seriously when encountering a large number of sequential tasks. LAMOL obtains competitive Intent ACC scores and even achieves the best EER score in the Blurry-domain setup, while it is far from reliable considering its poor performance in terms of JGA and BLEU. ToDCL achieves substantially better results than the other baselines by introducing a residual adapter for each task. Our MetaLTDS performs even better than ToDCL on most of the evaluation metrics. In addition, MetaLTDS is the lifelong TDS method with the highest efficiency, which requires much fewer parameters than baseline methods. This is a benefit of our adaptive masking learning method which reduces the redundancy of neurons.

We also demonstrate the middle states of the learned lifelong models. In Figure 2, we illustrate the average testing results (i.e., BLEU and JGA scores) over the first t tasks after learning the t-th task ($t \in [1, T]$) under the disjoint-domain setup. From the results, we can observe that the JGA scores of Finetune, EWC, AGEM and LAMOL drop sharply as more tasks arrive (i.e., $t = 5$). This indicates that these baseline methods suffer from catastrophic forgetting. Replay, ToDCL and MetaLTDS are resilient and keep relatively stable performance throughout the whole learning process. In particular, our MetaLTDS method exhibits better performance than Replay and ToDCL on all tasks.

Forward and Backward Knowledge Transfer

Following (Lopez-Paz and Ranzato 2017), we also evaluate the effectiveness of the lifelong dialogue systems for forward knowledge transfer (FWT) and backward knowledge transfer (BWT) respectively after learning all the tasks. The experimental results are demonstrated in Table 2. We observe that Finetune, LAMOL and Replay struggle to perform forward and backward knowledge transfer probably because they are incapable of handling the interference between the previously learned knowledge and new knowledge. The negative BWT indicates that these methods have forgotten some gained knowledge before learning subsequent new tasks. ToDCL shows no backward knowledge transfer through the whole learning process. This is consistent with our claim that the mechanism of adding adapters independently for each task may resolve the forgetting issue but ignore the knowledge transfer.

In contrast, our MetaLTDS method exhibits a promising ability to facilitate both forward and backward knowledge transfer. Since MetaLTDS is the only method that achieves positive BWT, we can deduce that MetaLTDS exploits the subsequent knowledge to benefit previous tasks. In addition, the high FWT score of MetaLTDS reveals its generalization ability.

Adaptive Mask Learning	✓	✓	✓	
Balanced Meta Learning	✓	✓		
Diverse Sampling	✓			
JGA	52.73	52.28	48.19	44.70
BLEU	18.15	17.56	14.81	11.23

Table 3: The ablation test results under the disjoint-domain setup in terms of JGA and BLEU.

Ablation Study

To investigate the effectiveness of different components in MetaLTDS, we also conduct an ablation study under the disjoint-domain setup. First, we substitute the diverse sampling strategy with a commonly used random selection method. Second, we remove balanced meta learning from MetaLTDS to analyze its contribution. Third, we discard adaptive mask learning such that no parameters could be reserved for a certain task. Table 3 reports the ablation test performance (i.e., JGA and BLEU) of the learned tasks.

From the results, we can make the observations that balanced meta learning and adaptive mask learning are two indispensable strategies in MetaLTDS. Concretely, JGA and BLEU scores drop shapely when removing adaptive mask learning. The reason is that the information on preceding tasks could remain intact via the masks, while the unmasked parameters could be used for future learning. In addition, balanced meta learning promotes the results by enabling the forward and backward knowledge transfer across the preserved knowledge in the masked neurons. It further equips the MetaLTDS method with the generalization ability to be better adapted to future tasks. Furthermore, the diverse sampling strategy also contributes to the performance.

Conclusion

In this paper, we propose a novel two-stage lifelong TDS to mitigate catastrophic forgetting and encourage knowledge transfer simultaneously, inspired by the learning process of humans. In the first stage, we learn task-specific masks that adaptively preserve the knowledge of each visited task to mitigate catastrophic forgetting. In the second stage, we devise a balanced meta learning strategy for both forward and backward knowledge transfer in the lifelong learning process. We conduct extensive experiments on 13 tasks with the disjoint-domain setup and 16 tasks with the blurry-domain setup. MetaLTDS outperforms the strong baselines in terms of both effectiveness and efficiency.

Appendices

Implementation Details We use a pre-trained GPT-2 (Radford et al. 2019) model as our backbone model, which is composed of 12 transformer layers and 768 hidden dimensions. We use AdamW for model optimization. In Met-

Disjoint-domain Setup			
Domains	Train	Dev	Test
TMB_restaurant	2000	250	250
TMB_food-ordering	2000	250	250
TMA_movie	2000	250	250
TMB_sport	2000	250	250
TMA_coffee	2000	250	250
TMA_pizza	2000	250	250
TMB_movie	2000	250	250
TMB_hotel	2000	250	250
TMB_flight	2000	250	250
TMA_uber	2000	250	250
TMA_auto	2000	250	250
TMB_music	2000	250	250
TMA_restaurant	2000	250	250

Blurry-domain Setup			
Domains	Train	Dev	Test
sgd_events, sgd_flights	3963	685	1199
sgd_events, sgd_rentalcars	4063	573	1213
sgd_homes, sgd_rentalcars	3087	549	1084
sgd_calendar, sgd_events	2194	301	649
sgd_rentalcars, sgd_travel	2231	308	659
sgd_calendar, sgd_ridesharing	1579	233	436
sgd_buses, sgd_travel	2062	345	634
sgd_events, sgd_music	1637	249	486
sgd_events, sgd_movies	1163	176	352
sgd_music, sgd_weather	1399	187	440
sgd_media, sgd_weather	1102	149	338
sgd_calendar, sgd_restaurants	1607	241	485
sgd_banks, sgd_media	626	98	188
sgd_flights, sgd_trains	768	102	229
sgd_alarm, sgd_homes	581	83	172
sgd_payment, sgd_restaurants	758	114	221

Table 4: Statistics of datasets in the experiments.

aLTDS, we set the learning rate to 3e-3 for the adaptive mask learning stage and 3e-4 for η (or α) for the meta learning stage. We set the learning rate to 6.25e-5 for Finetune, EWC, AGEM, LAMOL and Replay, and 6.25e-3 for ToDCL. The number of subsets H for meta-update is set to 2. In the diverse sampling, the number of augmentation times A is set to 3. For Replay and MetaLTDS, we set the memory size K to 200 under the disjoint-domain setup and 50 for the blurry-domain setup. To enhance the reliability and stability of the results, we run each method three times with randomly initialized parameters and report the averaged scores.

Data statistics We evaluate our MetaLTDS approach on a total of 29 tasks from three benchmark TDS datasets under two setups. For the disjoint-domain setup, we conduct experiments on Task-Master 2019 (TM19) and Task-Master 2020 (TM20) datasets (Byrne et al. 2019) with 13 tasks. For the blurry-domain setup, we conduct experiments on 16 multi-domain TDS tasks drawn from the Schema Guided Dialogue (SGD) dataset (Rastogi et al. 2020). In Table 4, we report the training, validation and testing sets of the disjoint-domain setup and the blurry-domain setup respectively.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (No. 2022YFF0902100), the National Natural Science Foundation of China (61906185, 62006062, 62176076).

References

Ahn, H.; Cha, S.; Lee, D.; and Moon, T. 2019. Uncertainty-based continual learning with adaptive regularization. In *NeurIPS*, 4392–4402.

Ba, J.; Kiros, J. R.; and Hinton, G. E. 2016. Layer Normalization. *ArXiv*, abs/1607.06450.

Bang, J.; Kim, H.; Yoo, Y.; Ha, J.-W.; and Choi, J. 2021. Rainbow Memory: Continual Learning with a Memory of Diverse Samples. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 8214–8223.

Biesialska, M.; Biesialska, K.; and Costa-jussà, M. R. 2020. Continual Lifelong Learning in Natural Language Processing: A Survey. In *Proceedings of the 28th International Conference on Computational Linguistics*, 6523–6541. Barcelona, Spain (Online): International Committee on Computational Linguistics.

Byrne, B.; Krishnamoorthi, K.; Sankar, C.; Neelakantan, A.; Goodrich, B.; Duckworth, D.; Yavuz, S.; Dubey, A.; Kim, K.-Y.; and Cedilnik, A. 2019. Taskmaster-1: Toward a Realistic and Diverse Dialog Dataset. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 4516–4525. Hong Kong, China: Association for Computational Linguistics.

Chaudhry, A.; Ranzato, M.; Rohrbach, M.; and Elhoseiny, M. 2019a. Efficient Lifelong Learning with A-GEM. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net.

Chaudhry, A.; Ranzato, M.; Rohrbach, M.; and Elhoseiny, M. 2019b. Efficient Lifelong Learning with A-GEM. *ArXiv*, abs/1812.00420.

de Masson d'Autume, C.; Ruder, S.; Kong, L.; and Yogatama, D. 2019. Episodic Memory in Lifelong Language Learning. *CoRR*, abs/1906.01076.

Gao, T.; Yao, X.; and Chen, D. 2021. SimCSE: Simple Contrastive Learning of Sentence Embeddings. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 6894–6910. Online and Punta Cana, Dominican Republic: Association for Computational Linguistics.

Geng, B.; Yuan, F.; Xu, Q.; Shen, Y.; Xu, R.; and Yang, M. 2021. Continual Learning for Task-oriented Dialogue System with Iterative Network Pruning, Expanding and Masking. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, 517–523. Online: Association for Computational Linguistics.

Goodfellow, I. J.; Mirza, M.; Da, X.; Courville, A. C.; and Bengio, Y. 2014. An Empirical Investigation of Catastrophic Forgeting in Gradient-Based Neural Networks. In Bengio, Y.; and LeCun, Y., eds., *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778.

He, W.; Dai, Y.; Yang, M.; Sun, J.; Huang, F.; Si, L.; and Li, Y. 2022a. Unified dialog model pre-training for task-oriented dialog understanding and generation. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 187–200.

He, W.; Dai, Y.; Zheng, Y.; Wu, Y.; Cao, Z.; Liu, D.; Jiang, P.; Yang, M.; Huang, F.; Si, L.; Sun, J.; and Li, Y. 2022b. GALAXY: A Generative Pre-trained Model for Task-Oriented Dialog with Semi-supervised Learning and Explicit Policy Injection. In *AAAI*.

Hosseini-Asl, E.; McCann, B.; Wu, C.-S.; Yavuz, S.; and Socher, R. 2020. A Simple Language Model for Task-Oriented Dialogue. In Larochelle, H.; Ranzato, M.; Hadsell, R.; Balcan, M.; and Lin, H., eds., *Advances in Neural Information Processing Systems*, volume 33, 20179–20191. Curran Associates, Inc.

Houlsby, N.; Giurgiu, A.; Jastrzebski, S.; Morrone, B.; De Laroussilhe, Q.; Gesmundo, A.; Attariyan, M.; and Gelly, S. 2019. Parameter-Efficient Transfer Learning for NLP. In Chaudhuri, K.; and Salakhutdinov, R., eds., *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, 2790–2799. PMLR.

Hung, C.-Y.; Tu, C.-H.; Wu, C.-E.; Chen, C.-H.; Chan, Y.-M.; and Chen, C.-S. 2019. Compacting, Picking and Growing for Unforgetting Continual Learning. In *Advances in Neural Information Processing Systems*, 13647–13657.

Ke, Z.; Liu, B.; Wang, H.; and Shu, L. 2021. Continual Learning with Knowledge Transfer for Sentiment Classification. In Hutter, F.; Kersting, K.; Lijffijt, J.; and Valera, I., eds., *Machine Learning and Knowledge Discovery in Databases*, 683–698. Cham: Springer International Publishing. ISBN 978-3-030-67664-3.

Kirkpatrick, J.; Pascanu, R.; Rabinowitz, N.; Veness, J.; Desjardins, G.; Rusu, A. A.; Milan, K.; Quan, J.; Ramalho, T.; Grabska-Barwinska, A.; et al. 2017. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 3521–3526.

Lee, S. 2017. Toward Continual Learning for Conversational Agents. *ArXiv*, abs/1712.09943.

Li, Z.; and Hoiem, D. 2018. Learning without Forgetting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(12): 2935–2947.

Lopez-Paz, D.; and Ranzato, M. 2017. Gradient Episodic Memory for Continual Learning. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, 6470–6479. Red Hook, NY, USA: Curran Associates Inc. ISBN 9781510860964.

Madotto, A.; Lin, Z.; Zhou, Z.; Moon, S.; Crook, P.; Liu, B.; Yu, Z.; Cho, E.; Fung, P.; and Wang, Z. 2021. Continual Learning in Task-Oriented Dialogue Systems. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 7452–7467. Online and Punta Cana, Dominican Republic: Association for Computational Linguistics.

Mallya, A.; and Lazebnik, S. 2018. Packnet: Adding multiple tasks to a single network by iterative pruning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7765–7773.

Mazumder, S.; and Liu, B. 2022. Continual Learning Dialogue Systems - Learning during Conversation. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '22, 3429–3432. New York, NY, USA: Association for Computing Machinery. ISBN 9781450387323.

McCloskey, M.; and Cohen, N. J. 1989. Catastrophic interference in connectionist networks: The sequential learning problem. In *Psychology of learning and motivation*, volume 24, 109–165. Elsevier.

Mermillod, M.; Bugaiska, A.; and Bonin, P. 2013. The stability-plasticity dilemma: investigating the continuum from catastrophic forgetting to age-limited learning effects. *Frontiers in Psychology*, 4.

Papineni, K.; Roukos, S.; Ward, T.; and Zhu, W.-J. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, 311–318.

Peng, B.; Li, C.; Li, J.; Shayandeh, S.; Liden, L.; and Gao, J. 2021. Soloist: Building Task Bots at Scale with Transfer Learning and Machine Teaching. *Transactions of the Association for Computational Linguistics*, 9: 807–824.

Qu, H.; Rahmani, H.; Xu, L.; Williams, B. M.; and Liu, J. 2021. Recent Advances of Continual Learning in Computer Vision: An Overview. *ArXiv*, abs/2109.11369.

Radford, A.; Wu, J.; Child, R.; Luan, D.; Amodei, D.; and Sutskever, I. 2019. Language Models are Unsupervised Multitask Learners. Technical report, OpenAI.

Rajasegaran, J.; Khan, S.; Hayat, M.; Khan, F. S.; and Shah, M. 2020. iTAML: An Incremental Task-Agnostic Meta-learning Approach. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Rastogi, A.; Zang, X.; Sunkara, S.; Gupta, R.; and Khaitan, P. 2020. Towards Scalable Multi-domain Conversational Agents: The Schema-Guided Dialogue Dataset. In *AAAI*.

Riemer, M.; Cases, I.; Ajemian, R.; Liu, M.; Rish, I.; Tu, Y.; and Tesauro, G. 2019. Learning to Learn without Forgetting By Maximizing Transfer and Minimizing Interference. *ArXiv*, abs/1810.11910.

Serra, J.; Suris, D.; Miron, M.; and Karatzoglou, A. 2018. Overcoming Catastrophic Forgetting with Hard Attention to the Task. In Dy, J.; and Krause, A., eds., *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, 4548–4557. PMLR.

Sun, F.-K.; Ho, C.-H.; and yi Lee, H. 2020. LAMOL: LAnguage MOdeling for Lifelong Language Learning. In *ICLR*.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L. u.; and Polosukhin, I. 2017. Attention is All you Need. In Guyon, I.; Luxburg, U. V.; Bengio, S.; Wallach, H.; Fergus, R.; Vishwanathan, S.; and Garnett, R., eds., *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Volpi, R.; Larlus, D.; and Rogez, G. 2021. Continual Adaptation of Visual Representations via Domain Randomization and Meta-learning. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 4441–4451.

Wen, T.-H.; Gašić, M.; Mrkšić, N.; Su, P.-H.; Vandyke, D.; and Young, S. 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*, 1711–1721. Lisbon, Portugal: Association for Computational Linguistics.

woo Koh, H.; Kim, D.; Ha, J.-W.; and Choi, J. 2021. Online Continual Learning on Class Incremental Blurry Task Configuration with Anytime Inference. *ArXiv*, abs/2110.10031.

Wu, C.-S.; Madotto, A.; Hosseini-Asl, E.; Xiong, C.; Socher, R.; and Fung, P. 2019. Transferable Multi-Domain State Generator for Task-Oriented Dialogue Systems. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 808–819. Florence, Italy: Association for Computational Linguistics.

Yan, S.; Xie, J.; and He, X. 2021. DER: Dynamically Expandable Representation for Class Incremental Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 3014–3023.