SELECTIVE TOKEN GENERATION FOR FEW-SHOT LANGUAGE MODELING

Anonymous authors
Paper under double-blind review

ABSTRACT

Natural language modeling with limited training data is challenging problem, and many algorithms make use of large-scale pretrained language models (PLMs) for this due to its great generalization ability. Among these transfer learning algorithms from PLMs, additive learning that incorporates a task-specific adapter on top of the fixed PLM has been popularly used to alleviate the severe overfitting problem in the few-shot setting. However, this added task-specific adapter is generally trained by maximum likelihood estimation that can easily suffer from the so-called exposure bias problem, especially in sequential text generation. Therefore, in this work, we develop a novel additive learning algorithm based on reinforcement learning (RL) for few-shot natural language generation (NLG) tasks. In particular, we propose to use a selective token generation between the transformer-based PLM and the task-specific adapter during both training and inference. This output token selection between the two generators allows the adapter to take into account only on the task-relevant parts in sequence generation, and therefore makes it more robust to overfitting as well as more stable in RL training. In addition, in order to obtain the complementary adapter from the PLM for each few-shot task, we exploit a separate selecting module that is also simultaneously trained using RL. Experimental results on various few-shot NLG tasks including data-to-text generation and text summarization demonstrate that the proposed selective token generation significantly outperforms the previous additive learning algorithms based on the PLMs.

1 INTRODUCTION

Natural language processing (NLP) have recently achieved great progress using advanced neural language models (Radford et al., 2018; Devlin et al., 2019; Yang et al., 2019; Liu et al., 2019; Clark et al., 2020; Raffel et al., 2019; Lan et al., 2019; Lewis et al., 2020). However, these neural models typically require large-scale training data with high-quality annotations for each individual task, and solving a new NLP task that has only a few examples is still challenging problem (Yin, 2020). Especially, natural language generation (NLG) with limited training data is an important yet more difficult task due to its fast adaptation of sequential prediction models in a wide range of applications including text summarization, question answering, data-to-text generation, machine translation, etc (Peng et al., 2020; Chen et al., 2020; Xu et al., 2021; Schick & Schütze, 2020; Chang et al., 2021; Radford et al., 2019; Lewis et al., 2020; Brown et al., 2020).

More recently, pretrained language models (PLMs) have shown great generalization ability when combined with large-scale data and big transformer-based models (Devlin et al., 2019; Radford et al., 2019; Lewis et al., 2020; Brown et al., 2020; Subramanyam Kalyan et al., 2021). Therefore, transfer learning from transformer PLMs has been popularly used for few-shot NLG tasks with promising results. In specific, the use of PLM for few-shot NLG can be categorized into three approaches: 1) prompt-based, 2) finetuning, and 3) additive learning. Prompt-based approaches encode a task description and task-specific examples as a natural language prompt for few-shot text generation (Radford et al., 2019; Brown et al., 2020; Zheng & Huang, 2021; Schick & Schütze, 2020; Li & Liang, 2021). While these approaches can take full advantage of the universal natural language understanding and generation capabilities of large-scale PLMs without further training of the main model, these have some limitations in dealing with a large domain shift from the pretraining corpus data, tuning suitable task-specific prompts, and covering an increased size of conditioning examples.

On the other hand, finetuning of the PLM is able to explicitly impart task-specific knowledge to the model and hence lift the above limitations (Ziegler et al., 2019; Xu et al., 2021; Chen et al., 2020). However, these finetuned models are prone to overfitting when only a small amount of training data is available. In order to alleviate such an overfitting problem, additive learning has been extensively exploited by incorporating task-specific adapters into the PLM (Zeldes et al., 2020; Stickland & Murray, 2019).

In general, task-specialized adapters for few-shot NLG are trained by maximum likelihood estimation (MLE). While MLE is efficient in learning, it suffers from the exposure bias problem due to the difference in the training and inference mechanisms (He et al., 2019), and this problem can be severe with limited training data. Reinforcement learning (RL) is capable of resolving this exposure bias problem by sequential output sampling during training (Ranzato et al., 2015; Keneshloo et al., 2019; Shi et al., 2021). Moreover, it allows to leverage the target-specific sequence-level objectives such as BLEU and ROUGE (Wu et al., 2018; Guo et al., 2021). However, in the task of NLG, the exponentially large space of output sequences restricts the use of RL since it leads to high variance and unstable training which is more serious in the few-shot setting. In this work, we develop a novel RL-based additive learning algorithm on the transformer-based PLM to overcome these shortcomings and to improve the performance of few-shot NLG.

In particular, we first convert the NLG task to the sequential token generation task based on the transformer language model, and then propose a selective token generation between the PLM and the task-specific adapter, during both RL-based training and inference. The proposed output token selection enables to not only explicitly maintain a general prior knowledge from the frozen PLM but also focus only on the task-relevant parts in sequence generation. In addition, in few-shot learning this partial token generation makes the task-specific adapter more resilient to overfitting and furthermore reduces the overall output space which leads to stable RL training. Here, in order to make the two token generators (policies) complement each other as well as to realize the robust output selection at the token level on the fly, we exploit a separate token-level policy selector. It is noted that both the policy selector and the task-specific adapter are simultaneously learned by the RL algorithm. Experimental results on various few-shot NLG tasks show that the proposed selective token generation outperforms the previous PLM-based additive learning algorithms with the comprehensive (non-selective) token generation.

Our main contributions can be summarized as follows.

- A novel selective token generation between the PLM and the task-specific adapter is proposed for transformer-based few-shot NLG.
- A separate selecting module is exploited to adaptively determine each output token in a sequence both at training and testing time.
- RL is applied to train both the policy selector and the task-specific adapter that is complementary to the PLM in text generation.
- An extensive empirical validation on few-shot NLG tasks demonstrates that the proposed selective token generation performs better in comparison to the previous PLM-based additive learning algorithms.

2 BACKGROUND

2.1 NATURAL LANGUAGE GENERATION

The goal of NLG is to generate a text sequence $\mathbf{y} = [y_0, ..., y_T]$ for a given task, where y_t is the tth output token from a vocabulary $\mathcal{V}$, and T is the output sequence length. For this generation, we aims to model the distribution of $\mathbf{y}$ that is autoregressively factorized as $p_\theta(\mathbf{y}) = \prod_{t=0}^{T} p_\theta(y_t|\mathbf{y}_{<t})$, where θ denotes the model parameters and $\mathbf{y}_{<t} = [y_0, ..., y_{t-1}]$. Here, the conditional distribution to sample a token for each step, $p_\theta(y_t|\mathbf{y}_{<t})$, is defined by the softmax function on the output logits $f_\theta(y_t|\mathbf{y}_{<t})$. Note that in general, the language generation is conditioned on input context according to a given task. Here, we encode the conditioning context by the same sequential model for generating an output sequence, and for simplicity we omit it. In this work, we utilize the autoregressive transformer for our generative model.

2.2 ADDITIVE LEARNING FOR FEW-SHOT GENERATION

To effectively leverage the general linguistic knowledge, θ is first initialized by the PLM parameters, θ_{LM}, for NLG. Given N task-specific training instances, $\mathcal{D} = \{\mathbf{y}^{n*}\}_{n=1}^{N}$, where $\mathbf{y}^{n*}$ is the nth ground-truth output sequence, directly finetuning θ_{LM} using $\mathcal{D}$ can incur the severe overfitting problem when N is small in the few-shot scenario. Therefore, we add the task-specific adapter, g_{θ_a} parameterized by θ_a, on top of the PLM, and optimize only θ_a (Zeldes et al., 2020; Stickland & Murray, 2019). In specific, we reformulate $f(\cdot|\mathbf{y}_{<t};\theta) = W^T h(\mathbf{y}_{<t};\theta_h)$ where $W \in \mathbb{R}^{H\times|\mathcal{V}|}$ and $h \in \mathbb{R}^H$ denote the weight matrix and the penultimate representations[1], respectively, and $\theta = \{W, \theta_h\}$. Then, we define the task-specific conditional distribution as follows:

$$p(y_t|\mathbf{y}_{<t};\theta_{LM},\theta_a) = softmax\Big({W_{LM}}^T h_{LM}(\mathbf{y}_{<t}) + {W_a}^T g\big(h_{LM}(\mathbf{y}_{<t});\theta_g\big)\Big), \tag{1}$$

where $h_{LM}(\mathbf{y}_{<t}) = h(\mathbf{y}_{<t};\theta_{h,LM})$ and $\theta_a = \{W_a, \theta_g\}$. Here, the summation of the PLM logits and the adapter logits is motivated by auxiliary training. It is noted that in our additive learning θ_a is updated while θ_{LM} is kept frozen. Hence, in the following we omit θ_{LM} such that $p_{\theta_a}(y_t|\mathbf{y}_{<t}) = p(y_t|\mathbf{y}_{<t};\theta_{LM},\theta_a)$ for simplicity.

2.3 MAXIMUM LIKELIHOOD ESTIMATION (MLE)

Given a small amount of training data $\mathcal{D} = \{\mathbf{y}^{n*}\}_{n=1}^{N}$, MLE optimizes θ by maximizing the data log-likelihood as follows:

$$\hat{\theta} = \arg\max_{\theta} \sum_{n=1}^{N}\sum_{t=0}^{T} \log p_\theta(y_t^{n*}|\mathbf{y}_{<t}^{n*}). \tag{2}$$

Here, the output token at each step is conditioned on not the previous sampled tokens from the current model but the previous ground-truth tokens $\mathbf{y}_{<t}^{n*}$. Namely, tokens are drawn from the data distribution during training, which is opposed to that tokens are drawn from the model distribution at test time. This discrepancy, also known as the exposure bias, leads that the errors will be accumulated along the generated sequence at test time since the model is biased to only perform well on the ground-truth history distribution. Especially, this bias problem can be more severe in the few-shot training. In addition, the token-level cross-entropy loss in MLE training is different from the sequence-level test metrics such as BLEU and ROUGE that are commonly used in the tasks of NLG.

2.4 REINFORCEMENT LEARNING (RL)

As an alternative to MLE, RL is able to overcome the exposure bias problem of MLE by sequence-level sampling from the model distribution during training (Ranzato et al., 2015). Also, RL can improve the performance by directly optimizing the evaluation metrics (Guo et al., 2021). In order to use RL for our additive learning, we reformulate our text generation as an RL problem: at each time step t, the agent takes the current state $\mathbf{s}_t = \mathbf{y}_{<t}$ as an input and performs an action a_t that outputs a token y_t by a policy $\pi_\theta(a_t|\mathbf{s}_t)$ corresponding to $p_\theta(y_t|\mathbf{y}_{<t})$. Then, the agent receives a reward $r_t = r(\mathbf{s}_t, a_t)$ and deterministically transitions to the next state $\mathbf{s}_{t+1}$. Here, note that the token-level intermediate reward $r_t = 0, \forall t < T$ when we use the delayed reward associated with the sequence-level evaluation metric between the two full sequences, $\mathbf{y}$ and $\mathbf{y}^*$. Let $\tau = \{(\mathbf{s}_t, a_t, r_t)\}_{t=0}^{T}$ be the trajectory generated by π_θ. The RL objective for the optimal agent is to maximize the expected sum of future discounted rewards:

$$J(\pi_\theta) = \mathbb{E}_{\tau\sim\pi_\theta}\Big[\sum_{t=0}^{T}\gamma^t r_t\Big], \tag{3}$$

where $\gamma \in [0, 1)$ is the discount factor.

Among a number of algorithms to approximately optimize the RL objective, we employ an actor-critic algorithm (Bahdanau et al., 2017) since it explicitly optimizes the policy network and it can

[1]For simplicity we consider h as not a sequence encoding function but the obtained representations.

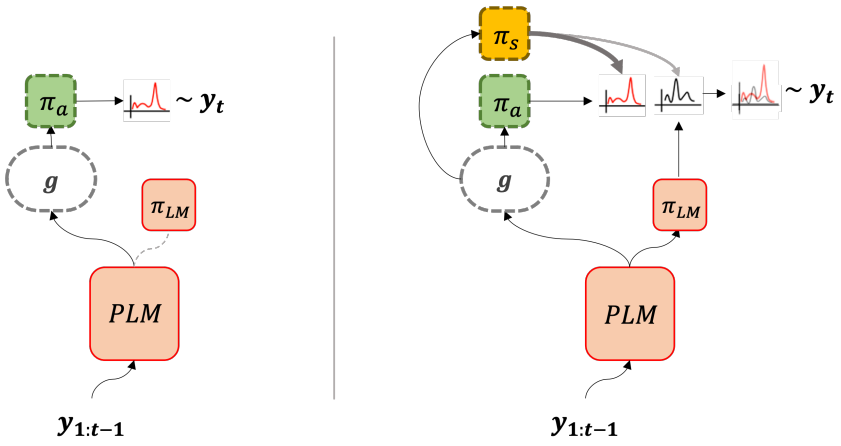

Figure 1: Text generation processes of Non-STG and STG are described. In the Non-STG model, every token is sampled from the task-specific policy π_a (**Left**). On the other hand, in the proposed STG model, each token is selectively sampled from either the PLM policy π_{LM} or the test-specific policy π_a where the selection is performed by the selection policy π_s (**Right**). Symbols with dashed line represent learnable models.

also alleviate the delayed reward problem. The actor-critic algorithm requires the additional critic network to estimate the value of a state, $V^\pi(\mathbf{s}_t) = \mathbb{E}_\pi[\sum_{t'=t}^{T} \gamma^{t'-t} r_{t'}|\mathbf{s}_t] = \sum_{a_t} \pi(a_t|\mathbf{s}_t) Q^\pi(\mathbf{s}_t, a_t)$ where the state-action value function $Q^\pi(\mathbf{s}_t, a_t) = \mathbb{E}_\pi[\sum_{t'=t}^{T} \gamma^{t'-t} r_{t'}|\mathbf{s}_t, a_t] = r_t + V^\pi(\mathbf{s}_{t+1})$. We use the following policy gradient loss to learn the policy parameters θ:

$$\mathcal{L} = -\sum_{t=0}^{T} A^{\pi_\theta}(\mathbf{s}_t, a_t) \log \pi_\theta(a_t|\mathbf{s}_t), \tag{4}$$

where $A^{\pi_\theta}(\mathbf{s}_t, a_t) = Q^{\pi_\theta}(\mathbf{s}_t, a_t) - V^{\pi_\theta}(\mathbf{s}_t)$ is the advantage function that quantifies how an action a_t is better than the average action in state $\mathbf{s}_t$.

In few-shot text generation, the extremely large action space ($|\mathcal{V}|^T$) as well as the small amount of training data often make it difficult to perform RL with degraded performances, even though we conduct the additive learning from the PLM. Therefore, in this work, we propose a selective token generation for improving the RL-based additive learning.

3 Selective Token Generation

Instead of generating all tokens in an output sequence from the single task-specific policy, $\pi_a = \pi_{\theta_a}(a_t|\mathbf{s}_t)$, at each time step t, we sample an output token y_t selectively from either the PLM policy $\pi_{LM} = \pi_{\theta_{LM}}(a_t|\mathbf{s}_t)$ or the task-specific policy π_a:

$$y_t = a_t \sim \big(\mathbb{1}_t[\pi_{LM} \text{ is selected}]\pi_{LM}(a_t|\mathbf{s}_t) + (1 - \mathbb{1}_t[\pi_{LM} \text{ is selected}])\pi_a(a_t|\mathbf{s}_t)\big), \tag{5}$$

where $\mathbb{1}_t[\cdot]$ is the indicator function (at t) that equals 1 if it is true and 0 otherwise. This output token selection allows to explicitly utilize a general linguistic knowledge from the PLM without catastrophic forgetting in few-shot learning. Also, the task-specific policy can focus on generating only the task-relevant parts, which enables more effective few-shot training with a reduced action space.

Now we need to determine how to select the proper policy at each step on the fly as well as to make the task-specific policy complementary to the PLM policy. For this, we exploit a separate token-level policy selector. The proposed policy selector $\pi_s(i_t|\mathbf{s}_t; \theta_s)$ with the parameters θ_s, where $i_t \in \{0, 1\}$, is an another policy that stochastically decides a policy to generate a_t for $\mathbf{s}_t$. Namely, a

token sample y_t is generated by the following process:

$$i_t \sim \pi_s(i_t|\mathbf{s}_t), \tag{6}$$

$$y_t = \begin{cases} a_t \sim \pi_{LM}(a_t|\mathbf{s}_t) & \text{if } i_t = 0, \\ a_t \sim \pi_a(a_t|\mathbf{s}_t) & \text{if } i_t = 1. \end{cases} \tag{7}$$

This process can be considered as a token generation from a hierarchical policy $\pi_h(a_t|\mathbf{s}_t;\theta_s,\theta_{LM},\theta_a)$ where the policy selector represents the upper-level prior for the preference of the low-level policy. Therefore, the value function of this hierarchical policy can be formulated as

$$\begin{aligned} V^{\pi_h}(\mathbf{s}_t) = \mathbb{E}_{\pi_h}[\sum_{t'=t}^{T}\gamma^{t'-t}r_{t'}|\mathbf{s}_t] &= \pi_s(i_t=0|\mathbf{s}_t)\sum_{a_t}\pi_{LM}(a_t|\mathbf{s}_t)Q^{\pi_h}(\mathbf{s}_t,a_t) \\ &\quad + \pi_s(i_t=1|\mathbf{s}_t)\sum_{a_t}\pi_a(a_t|\mathbf{s}_t)Q^{\pi_h}(\mathbf{s}_t,a_t), \end{aligned} \tag{8}$$

and $A^{\pi_h}(\mathbf{s}_t,a_t) = Q^{\pi_h}(\mathbf{s}_t,a_t) - V^{\pi_h}(\mathbf{s}_t)$. Here, it is noted that a single critic network is used for the hierarchical policy since i_t does not affect $\mathbf{s}_t$. Given a sample trajectory $\{(\mathbf{s}_t,i_t,a_t,r_t)\}_{t=0}^{T}$, the loss for optimizing θ_s and θ_a is

$$\begin{aligned} \mathcal{L} = -\sum_{t=0}^{T}A^{\pi_h}(\mathbf{s}_t,a_t)\Big[&\mathbb{1}[i_t=0]\big(\log\pi_{LM}(a_t|\mathbf{s}_t)+\log\pi_s(i_t|\mathbf{s}_t)\big) \\ &+\mathbb{1}[i_t=1]\big(\log\pi_a(a_t|\mathbf{s}_t)+\log\pi_s(i_t|\mathbf{s}_t)\big)\Big]. \end{aligned} \tag{9}$$

Similar to π_a, π_s makes use of the PLM representations and the task-specific adapter such that

$$\pi_s(i_t|\mathbf{s}_t;\theta_s) = softmax\Big(m\Big(g\big(h_{LM}(\mathbf{s}_t)\big);\theta_s\Big)\Big), \tag{10}$$

where m is the selector module that is implemented by a small neural network. Figure 1 depicts the overall text generation process by the proposed selective token generation (STG) in comparison to the previous non-selective token generation (Non-STG). Here, note that since all policies in STG share the same PLM representations, the increased computational cost by STG over Non-STG is negligible.

The use of the separated policy selector that is simultaneously trained with the task-specific policy allows the task-specific policy to be complementary to the PLM policy. Especially, this cooperative ensemble learning can be realized by our RL algorithm that performs sequential sampling from the model during training. In addition, the proposed STG model makes use of the PLM not at the feature level but the output distribution level in text generation. In our few-shot learning this is beneficial in explicitly retaining strong linguistic and world knowledge from the PLM.

4 EXPERIMENTS

In this section, we evaluate our method against baselines on Data-to-Text and Text Summarization tasks which are widely used in few-shot NLG. Since we claim the effectiveness of our method in additive learning, we compare our results against those obtained by other additive learning methods. We first explain baselines, implementation of our method, tasks for comparison and obtained results and then analyze the results in detail.

4.1 BASELINE

PLM. In our experiments, we assume that the PLM works to some extent for a given task. However, the naive PLM usually does not satisfy it for a new task unseen during training. Hence, we finetuned GPT[2] (Radford et al., 2019) with MLE for few epochs and used it as the PLM. Fine-tuning the PLM

[2]We experimented using GPT-2 with 345M parameters as the initial checkpoint. We follow the training details in the previous works (Peng et al., 2020; Khandelwal et al., 2019) for each task.

Table 1: Data-to-Text performance on FewShotWOZ dataset.

Model	Restaurant		Hotel		TV		Laptop	
	BLEU ↑	ERR ↓	BLEU ↑	ERR ↓	BLEU ↑	ERR ↓	BLEU ↑	ERR ↓
PLM	19.42	12.57	35.84	13.74	29.00	9.15	28.27	9.31
Non-STG-MLE	17.21	15.87	28.42	12.64	29.83	10.05	26.76	10.52
Non-STG-RL	18.01	11.98	36.72	12.64	28.66	9.19	28.59	9.21
NE(max)-MLE	14.12	15.27	31.32	14.29	28.23	10.21	26.93	10.02
NE(mix)-MLE	**25.27**	14.97	37.13	15.93	**32.85**	16.31	**32.91**	14.77
NE(max)-RL	15.20	11.68	32.68	16.48	28.91	9.24	28.66	9.51
NE(mix)-RL	24.10	19.16	38.07	18.68	32.84	18.06	32.53	17.14
STG	21.28	**10.78**	**38.09**	**11.54**	30.24	**9.03**	30.41	**8.91**

with MLE is most commonly used for task adaptation and thus it can also be a strong baseline. Note that this fine-tuning phase can be skipped when the advanced networks were used in the PLM such as External Encoders (Zeldes et al., 2020; Chen et al., 2020) or Adapters (Houlsby et al., 2019) which can cover the large domain shift.

Non-STG. This method stands for Non-Selective Token Generation which uses the above finetuned PLM as an encoder (frozen) and adapter (additional layer to be trained) followed by the encoder. We use two objectives, MLE and RL, for additive learning. These will be denoted as Non-STG-MLE and Non-STG-RL, respectively.

STG-Naive Ensemble. We believe that our proposed method encourages the task-specific policy (π_a) to complement the PLM's policy (π_{LM}) with a proper selection of the selector through joint training. To investigate this, we evaluate against two different naive ensembles of the policies, π_a trained from Non-STG and π_{LM} of the PLM. These ensemble schemes are the following:

- NE(max): $\pi_{max} = Max(\pi_a, \pi_{LM})$
- NE(mix): $\pi_{mix} = (\pi_a + \pi_{LM})/2$

4.2 Implementation

Here, we explain the implementation set-up of our method.

Additional Layer: The task-specific adapter g in Section 2.2 is implemented by using LSTM units[3] to encode the dynamics of the representation vector h_{LM}.

Selector. We use one or two $linear$ layers for m and the intermediate layer is activated by **ReLU** unit.

Reinforcement Learning. We employ Actor-Critic method (Konda & Tsitsiklis, 2000; Fedus et al., 2018) for RL. The agents (i.e. selector and generator) receive a reward after generating a sentence. Here, we use different reward functions according to tasks. We used BLEU score for (Peng et al., 2020) Data-to-Text and ROUGE-L for Text Summarization as the reward function.

Token Sampling. During the training, $i_t \in \{0, 1\} \sim \pi_s$ is sampled in advance and then we used π_{LM} of the PLM for $i_t = 0$ and task-specific policy π_a for $i_t = 1$ in token sampling. During the evaluation, we used an ensembled policy $\pi_h(\cdot) = \pi_s(i_t = 0)\pi_{LM}(\cdot) + \pi_s(i_t = 1)\pi_a(\cdot)$ in token sampling. We can use any encoding strategy such as beam search or sampling algorithm over the distribution. We used beam search with $width = 3$ for Text Summarization and $top_p = 0.9$ sampling for Data-to-Text.

4.3 Data-to-Text

Data-to-Text is a task that transforms structured data such as graphs or tables into natural language. Recent works (Mager et al., 2020; Peng et al., 2020; Kale, 2020) show that the PLM can be adapted successfully to this task by taking a serialized form of data as input without advanced models carefully designed to encode the structured data. Here, we experiment on FewShotWOZ (Peng et al.,

[3]We found that the methods using additional layers with MLP can not outperform the PLM.

2020) dataset[4]. There are 50 training instances for each topic and 129, 78, 1379, and 680 testing instances for Restaurant, Hotel, Laptop, and TV respectively. The models are evaluated by measuring fluency and informativeness using BLEU score and ERR (slot ERror Rate), respectively.

4.4 TEXT SUMMARIZATION

Here, we consider the problem of abstractive summarization: models aim to generate its summary (the "target" text) for a given piece of "source" text such that its meaning is intact. The model has to learn 1) building a semantic form internally from the source text and 2) converting the semantic form into a text. Khandelwal et al. (2019) showed that the PLM trained with only 1% of samples of the dataset shows comparable results to existing advanced models. Following this concept, we randomly sample various sizes of (0.5%, 1%, and 2%) subset data from CNN/Daily Mail (See et al., 2017). We collect three different subsets for each size with different random seeds. We evaluate the models against these nine datasets. ROUGE (Lin, 2004) is commonly employed to evaluate n-grams recall of the summaries with gold references. ROUGE-1, ROUGE-2, and ROUGE-L (measured based on longest common sub-sequence) results are reported in Table. 2.

4.5 RESULT

We compare experiment results on various tasks in this subsection. In the most experimental results, we found that additive learning can be used to improve performance. However, they do not always guarantee a performance improvement. For example, the ERR score of the PLM on *Laptop* shows a better result except for STG and NE(mix)-RL (See Table 1).

In Data-to-Text task, we consider two metrics, BLEU which measures the fluency of sentence and ERR which measures the rate of missing information from the given structured data. Note that Data-to-Text has a very small amount of training data (50) which means that it is easy to be overfitted in training. We can observe worse BLEU (3/4) and ERR (3/4) for Non-STG-MLE, worse BLEU (3/4) and better ERR (3/4) for Non-STG-RL, worse BLEU (4/4) and ERR (4/4) for NE(max)-MLE, better BLEU (4/4) and worse ERR (4/4) for NE(mix)-MLE, worse BLEU (3/4) and ERR (3/4) for NE(max)-RL, better BLEU (4/4) worse ERR(4/4) for NE(mix)-RL and better BLEU (4/4) and ERR (4/4) for STG as shown in Table 1. While other methods show some trade-off between BLEU and ERR, only our method (STG) shows improvements on both BLEU and ERR for all topics in the dataset.

In Text Summarization (a problem of long-sequence generation), our method shows significantly larger gains than Non-STG-MLE, Non-STG-RL and their naive ensembles with the PLM in every score metric and training data size as shown in Table 2. We expect that this result is induced from the complementary behavior between MLE-based and RL-base method which takes the advantages of each other. We will discuss details of it in Section 4.6.

4.6 WHAT MAKES STG BETTER?

As we mentioned in Section 4.5, we expect that our performance improvement comes from complementary behaviors between the PLM and RL-based training. We explain each of them in this section.

Knowledge Preservation. Since Non-STG-MLE and Non-STG-RL do not freeze the PLM, they may lose the general knowledge of the PLM trained from a large amount of data which can be helpful in the task as the training proceeds in a task-specific dataset. This phenomenon will be much more severe when there is a small amount of training dataset since it doesn't contain enough knowledge of a task. However, our method is fully accessible to this general knowledge of the PLM since we freeze it and dynamically use it with the selector's policy during a sequence generation. Hence, we claim that our method has a potential to perform better with its more general knowledge.

Taking Advantages of RL. As we mentioned in Section 1, RL in NLG has many strengths: 1) It is free from exposure bias which can seriously degrade the performance with accumulated error, 2) it is generally more resilient to overfitting since it experiences diverse sequences during training even

[4]We only tested on the topics (Restaurant, Hotel, Laptop, TV) because the evaluator has not published yet for the other topics (Taxi, Attraction, Train). `https://github.com/pengbaolin/SC-GPT`

Table 2: Averaged performance gain against PLM for Text Summarization on various size of few-shot subset data of CNN/DM.

Model	0.5%			1%			2%		
	R1	R2	RL	R1	R2	RL	R1	R2	RL
PLM	30.19	11.27	21.21	31.03	11.56	21.62	33.05	12.96	23.39
Non-STG-MLE	0.14	0.05	-0.01	0.22	0.07	0.10	0.14	0.02	0.00
Non-STG-RL	0.15	0.07	0.02	0.21	0.08	0.13	0.17	0.04	0.01
NE(max)-MLE	0.14	0.04	0.00	0.22	0.07	0.09	0.14	0.03	0.01
NE(mix)-MLE	0.13	0.04	0.02	0.17	0.04	0.06	0.06	0.03	0.02
NE(max)-RL	0.18	0.08	0.06	0.24	0.10	0.15	0.16	0.04	0.02
NE(mix)-RL	0.09	0.03	0.02	0.15	0.06	0.07	0.09	0.05	0.03
STG	**0.28**	0.09	**0.15**	**1.00**	**0.40**	**0.61**	**0.40**	**0.18**	**0.27**

though the amount of training data is small and 3) it can leverage target-specific objectives such as BLEU and ROUGE. Hence, one can notice that RL is a promising method if it is successfully applied. However, RL has some limitation with its application: 1) it suffers from exponentially large search space $|\mathcal{V}|^T$ and 2) it requires considerably many steps of training since it generally learns from the uniform random policy which tries obviously meaningless actions. Our method resolves these issues and successfully applies RL as explained in the next paragraph.

Resolving Disadvantages of RL. We resolve the issues in the application of RL using the PLM. We resolve the first issue with our selective structure with the PLM. Since frozen PLM chooses a token when it is selected, the search space of the generator is approximately decreased from $|\mathcal{V}|^T$ to $|\mathcal{V}|^{T-\overline{T}_{PLM}}$ where $\overline{T}_{PLM}$ is average length of sequences generated by PLM. Also, We resolve the second issue with our logit-summation structure of task-specific policy as described in Eq. 1. Since the adapter logits are nearly zero before training is advanced, it satisfies $\pi_a \approx \pi_{LM}$ which lets the generator start learning from the finetuned policy of PLM, not a uniform random policy.

5 Related Work

Recently, prompt-based in-context learning with an extremely large transformer-based PLM shows impressive few-shot generation performances (Radford et al., 2019; Brown et al., 2020). Schick & Schütze (2020) propose manually designed natural language prompts for improved few-shot text summarization and headline generation. Elsahar et al. (2018) conduct zero-shot learning for question generation from knowledge graphs, however they require a large amount of in-domain training data for their transfer learning. Chen et al. (2020) directly finetune the pretrained GPT-2 with a small amount of serialized attribute-value pairs for table-to-text generation. Gong et al. (2020) further apply multiple tasks to effectively leverage the structured information of tables. In contrast to these approaches, our proposed method utilizes RL-based additive learning for few-shot text generation.

Applying RL for text generation has been widely used to mitigate the exposure bias problem of MLE as well as to directly optimize task-relevant evaluation metrics. Ranzato et al. (2015) use the REINFORCE algorithm for text summarization and machine translation while Bahdanau et al. (2017) use the actor-critic algorithm for machine translation. However, they require pretraining using MLE. Ding & Soricut (2017) propose softmax policy gradient to remove the MLE-based pretraining. However, it requires various techniques for effective training. Tan et al. (2018) propose an entropy-regularized policy optimization that subsumes many of the previous training algorithms. Our proposed method is different from these methods in that we apply RL for more difficult few-shot generative modeling.

Various methods take into account the RL tasks with large action spaces like NLG. Dulac-Arnold et al. (2015) consider only actions in a cluster around the latent state of action obtained from a given state. Chandak et al. (2019) define the action embedding as a distribution with semantic of action and use a deterministic policy to take an action. Even-Dar et al. (2003); Zahavy et al. (2018) devise a method of incorporating the process of directly removing unnecessary actions according to the state in the RL problem. Guo et al. (2021) propose to use soft Q-learning and path consistency learning to combine off- and on-policy updates. Unlike these approaches, we use the hierarchical policy that reduces the sequential action space.

6 CONCLUSION

In this work, we propose to exploit a selective token generation between the pretrained language model and the task-specific adapter with RL-based additive learning for the tasks of few-shot natural language generation. In particular, we devise a trainable policy selector at the token level and jointly learn it with the task-specific policy. The proposed policy selector and RL algorithm make the two policies complementary each other and lead to robust few-shot generative modeling. Experimental results on various tasks of few-shot text generation show that the proposed selective token generation along with RL-based additive learning consistently improves the performances with less overfitting. For future work, we will investigate more general ensemble learning for few-shot learning and perform more study on the architectures of both the adapter and selector.

REFERENCES

Dzmitry Bahdanau, Philemon Brakel, Kelvin Xu, Anirudh Goyal, Ryan Lowe, Joelle Pineau, Aaron Courville, and Yoshua Bengio. An actor-critic algorithm for sequence prediction. *International Conference on Learning Representations*, 2017. URL https://arxiv.org/abs/1607.07086.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*, 2020.

Yash Chandak, Georgios Theocharous, James Kostas, Scott Jordan, and Philip Thomas. Learning action representations for reinforcement learning. In *International Conference on Machine Learning*, pp. 941–950. PMLR, 2019.

Ernie Chang, Xiaoyu Shen, Alex Marin, and Vera Demberg. The selectgen challenge: Finding the best training samples for few-shot neural text generation. *arXiv preprint arXiv:2108.06614*, 2021.

Zhiyu Chen, Harini Eavani, Wenhu Chen, Yinyin Liu, and William Yang Wang. Few-shot NLG with pre-trained language model. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 183–190, Online, July 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.18. URL https://aclanthology.org/2020.acl-main.18.

Kevin Clark, Minh-Thang Luong, Quoc V. Le, and Christopher D. Manning. Electra: Pre-training text encoders as discriminators rather than generators. *International Conference on Learning Representations*, 2020.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pp. 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics. doi: 10.18653/v1/N19-1423. URL https://www.aclweb.org/anthology/N19-1423.

Nan Ding and Radu Soricut. Cold-start reinforcement learning with softmax policy gradient. *arXiv preprint arXiv:1709.09346*, 2017.

Gabriel Dulac-Arnold, Richard Evans, Hado van Hasselt, Peter Sunehag, Timothy Lillicrap, Jonathan Hunt, Timothy Mann, Theophane Weber, Thomas Degris, and Ben Coppin. Deep reinforcement learning in large discrete action spaces. *arXiv preprint arXiv:1512.07679*, 2015.

Hady Elsahar, Christophe Gravier, and Frederique Laforest. Zero-shot question generation from knowledge graphs for unseen predicates and entity types. *arXiv preprint arXiv:1802.06842*, 2018.

Eyal Even-Dar, Shie Mannor, and Yishay Mansour. Action elimination and stopping conditions for reinforcement learning. In *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, pp. 162–169, 2003.

William Fedus, Ian Goodfellow, and Andrew M Dai. Maskgan: better text generation via filling in the_. *International Conference on Learning Representations*, 2018.

Heng Gong, Yawei Sun, Xiaocheng Feng, Bing Qin, Wei Bi, Xiaojiang Liu, and Ting Liu. Tablegpt: Few-shot table-to-text generation with table structure reconstruction and content matching. In *International Conference on Computational Linguistics*, pp. 1978—-1988, 2020.

Han Guo, Bowen Tan, Zhengzhong Liu, Eric P. Xing, and Zhiting Hu. Text generation with efficient (soft) q-learning. *arXiv preprint arXiv:2106.07704*, 2021.

Tianxing He, Jingzhao Zhang, Zhiming Zhou, and James Glass. Exposure bias versus self-recovery: Are distortions really incremental for autoregressive text generation? *arXiv preprint arXiv:1905.10617*, 2019.

Neil Houlsby, Andrei Giurgiu, Stanislaw Jastrzebski, Bruna Morrone, Quentin De Laroussilhe, Andrea Gesmundo, Mona Attariyan, and Sylvain Gelly. Parameter-efficient transfer learning for nlp. In *International Conference on Machine Learning*, pp. 2790–2799. PMLR, 2019.

Mihir Kale. Text-to-text pre-training for data-to-text tasks. *arXiv preprint arXiv:2005.10433*, 2020.

Yaser Keneshloo, Tian Shi, Naren Ramakrishnan, and Chandan K Reddy. Deep reinforcement learning for sequence-to-sequence models. *IEEE transactions on neural networks and learning systems*, 31(7):2469–2489, 2019.

Urvashi Khandelwal, Kevin Clark, Dan Jurafsky, and Lukasz Kaiser. Sample efficient text summarization using a single pre-trained transformer. *arXiv preprint arXiv:1905.08836*, 2019.

Vijay R Konda and John N Tsitsiklis. Actor-critic algorithms. In *Advances in neural information processing systems*, pp. 1008–1014. Citeseer, 2000.

Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. Albert: A lite bert for self-supervised learning of language representations. *International Conference on Learning Representations*, 2019.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 7871–7880, 2020.

Xiang Lisa Li and Percy Liang. Prefix-tuning: Optimizing continuous prompts for generation. *arXiv preprint arXiv:2101.00190*, 2021.

Chin-Yew Lin. ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pp. 74–81, Barcelona, Spain, July 2004. Association for Computational Linguistics. URL `https://aclanthology.org/W04-1013`.

Yinhan Liu, Naman Goyal Myle Ott, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.

Manuel Mager, Ramón Fernandez Astudillo, Tahira Naseem, Md Arafat Sultan, Young-Suk Lee, Radu Florian, and Salim Roukos. Gpt-too: A language-model-first approach for amr-to-text generation. *arXiv preprint arXiv:2005.09123*, 2020.

Baolin Peng, Chenguang Zhu, Chunyuan Li, Xiujun Li, Jinchao Li, Michael Zeng, and Jianfeng Gao. Few-shot natural language generation for task-oriented dialog. *arXiv preprint arXiv:2002.12328*, 2020.

A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever. Improving language understanding by generative pre-training. 2018.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9, 2019.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. *arXiv preprint arXiv:1910.10683*, 2019.

Marc'Aurelio Ranzato, Sumit Chopra, Michael Auli, and Wojciech Zaremba. Sequence level training with recurrent neural networks. *arXiv preprint arXiv:1511.06732*, 2015.

Timo Schick and Hinrich Schütze. Few-shot text generation with pattern-exploiting training. *arXiv preprint arXiv:2012.11926*, 2020.

Abigail See, Peter J Liu, and Christopher D Manning. Get to the point: Summarization with pointer-generator networks. *arXiv preprint arXiv:1704.04368*, 2017.

Tian Shi, Yaser Keneshloo, Naren Ramakrishnan, and Chandan K Reddy. Neural abstractive text summarization with sequence-to-sequence models. *ACM Transactions on Data Science*, 2(1): 1–37, 2021.

Asa Cooper Stickland and Iain Murray. Bert and pals: Projected attention layers for efficient adaptation in multi-task learning. In *International Conference on Machine Learning*, pp. 5986–5995. PMLR, 2019.

Katikapalli Subramanyam Kalyan, Ajit Rajasekharan, and Sivanesan Sangeetha. Ammus : A survey of transformer-based pretrained models in natural language processing. *arXiv preprint arXiv:2108.05542*, 2021.

Bowen Tan, Zhiting Hu, Zichao Yang, Ruslan Salakhutdinov, and Eric Xing. Connecting the dots between mle and rl for sequence prediction. *arXiv preprint arXiv:1811.09740*, 2018.

Lijun Wu, Fei Tian, Tao Qin, Jianhuang Lai, and Tie-Yan Liu. A study of reinforcement learning for neural machine translation. *arXiv preprint arXiv:1808.08866*, 2018.

Xinnuo Xu, Guoyin Wang, Young-Bum Kim, and Sungjin Lee. Augnlg: Few-shot natural language generation using self-trained data augmentation. *arXiv preprint arXiv:2106.05589*, 2021.

Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Russ R. Salakhutdinov, and Quoc V. Le. Xlnet: Generalized autoregressive pretraining for language understanding. In *Advances in neural information processing systems*, pp. 5753–5763, 2019.

Wenpeng Yin. Meta-learning for few-shot natural language processing: A survey. *arXiv preprint arXiv:2007.09604*, 2020.

Tom Zahavy, Matan Haroush, Nadav Merlis, Daniel J Mankowitz, and Shie Mannor. Learn what not to learn: Action elimination with deep reinforcement learning. *arXiv preprint arXiv:1809.02121*, 2018.

Yoel Zeldes, Dan Padnos, Or Sharir, and Barak Peleg. Technical report: Auxiliary tuning and its application to conditional text generation. *arXiv preprint arXiv:2006.16823*, 2020.

Chujie Zheng and Minlie Huang. Exploring prompt-based few-shot learning for grounded dialog generation. *arXiv preprint arXiv:2109.06513*, 2021.

Daniel M. Ziegler, Nisan Stiennon, Jeffrey Wu, Tom B. Brown, Alec Radford, Dario Amodei, Paul Christiano, and Geoffrey Irving. Fine-tuning language models from human preferences. *arXiv preprint arXiv:1909.08593*, 2019.

A TRAINING SETTINGS

Table 3: **Hyper-parameters used for experiments**

Hyper-parameter	**Summarization**	**Data-to-Text**
Num layer	2	
RNN hidden size	512	256
γ	1	
Optimizer	AdamW	
Betas	(0.9, 0.999)	
Eps	1e-8	
Learning rate	2e-5	
Pre-train epoch (Non-STG-RL)	1	0
Validation	500	50
Train epochs	16 (0.5%), 8 (1%), 4 (2%)	30
Batch size	16	10

A.1 GENERATED SENTENCE EXAMPLES

Table 4: Data-to-Text generation examples from TV of FewShotWOZ dataset. The tokens colored in red are the selected tokens from the additional policy.

Data	inform(name=typhon 48; type=television; ecorating=a+; pricerange=cheap; hdmiport=4)
Ground-truth	the typhon 48 television has a+ eco rating , 4 hdmi ports and is in the cheap price range .
NE(mix)-MLE	the typhon 48 television is in the cheap price range and 4 hdmi ports.
STG	the typhon 48 television has 4 hdmi ports, a eco rating, and is in the cheap price range.
Data	inform(name=hesperus 33; type=television; resolution=720p; family=w2; screensize=32 inch)
Ground-truth	hesperus 33 is a 720p television in the w2 product family with a 32 inch screen size .
NE(mix)-MLE	the hesperus 33 television resolves 720p, has a 32 inch screen, and is in the w2 product family.
STG	the hesperus 33 television is in the w2 product family, has a 32 inch screen, and uses a 720p resolution.
Data	recommend(name=hymenaios 11;type=television; screensizerange=small; hdmiport=1; accessories=remote control)
Ground-truth	hymenaios 11 is a small screen size television with 1 hdmi ports , and a remote control .
NE(mix)-MLE	the hymenaios 11 television has 1 hdmi port and 1 remote control.
STG	the hymenaios 11 television is a nice television in the small screen size range of 1 hdmi port and 1 remote control.
Data	?compare(name=aristaeus 59; screensizerange=large; resolution=1080p; name=charon 61; screensizerange=medium; resolution=720p)
Ground-truth	please tell us which television you prefer: aristaeus 59 with large screen and 1080p resolution , or charon 61 with medium screen and 720p resolution ?
NE(mix)-MLE	the aristaeus 59 is a 1080p television in the medium screen size range and a 720p resolution.
STG	the aristaeus 59 features a large screen size and a medium resolution. it has a 1080p and a charon 61 that has a medium screen size and a 720p resolution.

Table 5: Example summaries of the baselines and proposed model on the CNN/DailyMail dataset

Article	-lrb- cnn -rrb- – three runners died sunday during the detroit free press/flagstar marathon in detroit, michigan, police told cnn. an emt vehicle is at the scene sunday in detroit after three runners collapsed at a marathon. all three deaths occurred between 9 and 9:20 a.m. et, second deputy chief john roach said. a man in his 60s fell and hit his head, roach said. the cause of the fall was unknown. the man was transported to detroit receiving hospital, where he was pronounced dead. two other men, ages 36 and 26, also collapsed during the race and were pronounced dead at the hospital, roach said. all three collapsed near the end of the race, he said. witnesses describe scene " the weather at the time was overcast, roach said, with temperatures in the low 40s. [...]
Ground-truth	second deputy chief john roach : all three deaths occurred between 9 and 9:20 a.m. man in his 60s fell hit his head ; two men others , ages 36 and 26 , collapsed . race was detroit free press/flagstar marathon in detroit , michigan .
PLM	three runners collapsed at a marathon in detroit , police say . the cause of the fall is unknown .
Non-STG-MLE	three runners collapsed at a marathon sunday , police say . the cause of the fall is unknown , police say .
Non-STG-RL	three runners collapsed at a marathon sunday , police say . the cause of the fall is unknown , police say .
STI	three runners collapsed at a marathon in detroit , michigan . all three deaths occurred between 9 and 9:20 a.m. et . a man in his 60s fell and hit his head , police say .
Article	london, england -lrb- cnn -rrb- – up to 1,000 human rights campaigners demonstrated saturday in front of no. 10 downing street, the official residence of british prime minister gordon brown, calling on the british government to demand that full democracy be restored in pakistan. jemima khan, center, ex-wife for former pakistani cricket star imran khan, joins protesters in london. protesters waved placards and chanted in support of the resignation of pakistani president pervez musharraf, a week after he imposed a state of emergency in the country. the crowd of demonstrators massed behind barriers and included jemima khan, the ex-wife of former pakistani cricket star turned politician imran khan. the demonstrators carried placards saying " free the innocent " and " end musharraf's regime " and waved pakistani flags. imran khan, who heads the the movement for justice party, has been under house arrest since the emergency declaration. his ex-wife delivered a petition to a doorman at downing street, calling on britain to use its influence to ensure that all institutions are in place well in advance of pakistani elections originally scheduled for early next year. the petition also [...]
Ground-truth	human rights campaigners demonstrate in front of no . 10 downing street . protests urged uk government to demand full democracy restored in pakistan . cricketer turned politician imran khan ś ex wife jemima among protesters.
PLM	pakistani president pervez musharraf has been under house arrest since the emergency declaration . his ex-wife delivered a petition to a doorman at downing street .
Non-STG-MLE	pakistani president pervez musharraf has imposed a state of emergency in the country . he has been under house arrest since the emergency declaration .
Non-STG-RL	pakistani president pervez musharraf has imposed a state of emergency in the country . he has imposed a state of emergency in the country since last week .
STG	the official residence of british prime minister gordon brown . hundreds of protesters demonstrate in front of no . 10 downing street . the petition calls for full democracy in pakistan .

Article	-lrb- cnn -rrb- spoiler alert! it's not just women getting cloned. that was the big twist at the end of " orphan black's " second season. the kickoff to the new season leads the list of six things to watch in the week ahead. 1. " orphan black, " 9 p.m. et, saturday, april 18, bbc america. the cloning cult sci-fi series remains one of the most critically acclaimed shows on tv, thanks in large part to the performance of tatiana maslany, who has taken on at least six roles on the show so far, including a newly introduced transgender clone. maslany told reporters this week that we can expect even more impressive scenes with multiple clones. " we like to push the boundaries of what we're able to do and the limits of those clone scenes, " she said. " so, yes, you 'll definitely see more complex clone work this season and that's just because we're getting more comfortable with the technology and we're excited by getting to sort of further complicate things. " and the introduction of a group of male clones will certainly increase the suspense. [...]
Ground-truth	critically acclaimed series " orphan black " returns . " turn : washington ś spies " starts a second season . " game of thrones " is back for season five.
PLM	series returns with a new subtitle , " washington ś spies , " and a new monday night time slot . series star jamie bell told cnn what we can expect in the second season .
Non-STG-MLE	órphan blackŕeturns with a new subtitle and a new monday night time slot . series star tatiana maslany is blown away by the response to the series so far .
Non-STG-RL	órphan blackŕeturns with a new subtitle and a new monday night time slot . series star jamie bell says the series will bring more recognition to the culper spy ring .
STG	the new season of " orphan black " premieres on saturday the 18th . " game of thrones " returns for a fifth season .
Article	energy drinks can pose a real problem for children, according to a new study from the american heart association. researchers found that 40 % of the 5,156 calls to poison centers for " energy drink exposure " involved children under age 6. in most of the cases, the parents didn't know the children had gotten hold of an energy drink. many of the calls reported the children were experiencing serious side effects, such as an abnormal heart rhythm, or they were having a seizure. the study is being presented at the american heart association's scientific sessions this week. study : caffeine affects teen boys more. study author dr. steven lipshultz has handled cases involving children who became sick after consuming energy drinks. while studies about the impact of caffeine on children are limited, lipshultz, a professor and chairman of pediatrics at wayne state university, said adolescents can experience serious problems after drinking only 100 mg of caffeine. [...]
Ground-truth	40 % of " energy drink exposure " calls to poison control centers are for children under 6. study authors believe the number is under-counted . the fda does not set a caffeine limit for children.
PLM	drug calls to poison centers for " energy drink exposure " in most cases , the parents did n't know the children had gotten hold of an energy drink . many of the calls reported the children were experiencing serious side effects , such as an abnormal heart rhythm .
Non-STG-MLE	drug calls to poison centers for " energy drink exposure " in most cases , the parents did n't know the children had gotten hold of an energy drink . many of the calls reported the children were experiencing serious side effects , such as an abnormal heart rhythm .
Non-STG-RL	" this is a very concerning finding , " dr. laurence sperling says . the american academy of pediatrics recommends children consume no caffeine .
STG	drug calls to poison centers for " energy drink exposure " nearly 40 % of calls to poison centers for " energy drink exposure " involved children under age 6. study : caffeine affects teens more .