Selective Token Generation for Few-shot Natural Language Generation

[1]Daejin Jo, [1]Taewhan Kwon, [2]Eun-Sol Kim, and [1]Sungwoong Kim

[1]Kakao Brain

[2]Hanyang University

Background

- **Few-shot NLG**
  - Adaptation of **Natural Language Generation** (NLG) tasks when **only a small amount of training data** is available.

Background

- **Few-shot NLG**
  - Adaptation of **Natural Language Generation** (NLG) tasks when **only a small amount of training data** is available.

- **Pretrained Language Model** (PLM) transfer methods
  1. Zero-shot or in-context few-shot learning (e.g. GPT)
     - Limited for novel tasks having large domain shift
     - Limited in covering an increased size of conditioning examples
  2. Fine-tuning PLM
     - Prone to overfitting
  3. **Additive learning** based on **task-specific adapter**
     - This approach can alleviate above issues.

Objectives of additive learning

- In general, task-specific adapters are trained by maximum likelihood estimation (MLE) or reinforcement learning (RL).
- MLE is usually chosen due to its efficiency but suffers from exposure bias.
- RL resolves exposure bias but challenging due to the training instability by exponentially large space of output sequences in NLG.

Objectives of additive learning

- In general, task-specific adapters are trained by maximum likelihood estimation (MLE) or reinforcement learning (RL).
- MLE is usually chosen due to its efficiency but suffers from exposure bias.
- RL resolves exposure bias but challenging due to the training instability by exponentially large space of output sequences in NLG.

→ *Existing additive learning produces* ***the whole sequence by its own adapter***

→ *This is a fundamental limitation in maintaining the knowledge of PLM*

Case study - QA

Passage	three types of conflicts are : 1. intrapersonal conflicts , 2. interpersonal conflicts and 3. unconscious conflicts . the word conflict has been derived from a latin word “conflicts” which means “strike two things at the same time” . conflict is [1)]**an opposition or a tug-of-war between contradictory impulses .** according to colman "a conflict is [2)]**the anticipated frustration entailed in the choice of either alternative**".
Query	conflict definition psychology
Ground-truth	the anticipated frustration entailed in the choice of either alternative.

1) General meaning of conflict

2) Psychological meaning of conflict

Case study - QA

Passage	three types of conflicts are : 1. intrapersonal conflicts , 2. interpersonal conflicts and 3. unconscious conflicts . the word conflict has been derived from a latin word "conflicts" which means "strike two things at the same time" . conflict is 1)**an opposition or a tug-of-war between contradictory impulses .** according to colman "a conflict is 2)**the anticipated frustration entailed in the choice of either alternative**".
Query	conflict definition psychology
Ground-truth	the anticipated frustration entailed in the choice of either alternative.

- Without the knowledge of who *Colman*[1] is, it can be hard to answer since the word psychology in the query does not appear in the passage.

[1]A psychologist, https://en.wikipedia.org/wiki/Peter_T._Coleman_(academic)

Case study - QA

Passage	three types of conflicts are : 1. intrapersonal conflicts , 2. interpersonal conflicts and 3. unconscious conflicts . the word conflict has been derived from a latin word "conflicts" which means "strike two things at the same time" . conflict is [1)]**an opposition or a tug-of-war between contradictory impulses .** according to colman "a conflict is [2)]**the anticipated frustration entailed in the choice of either alternative**".
Query	conflict definition psychology
Ground-truth	the anticipated frustration entailed in the choice of either alternative.
PLM	conflict definition psychology.

→ Lack of domain adaptation

- PLM repeats the given query as its generated answer due to the lack of domain adaptation.

Case study - QA

Passage	three types of conflicts are : 1. intrapersonal conflicts , 2. interpersonal conflicts and 3. unconscious conflicts . the word conflict has been derived from a latin word “conflicts” which means “strike two things at the same time” . conflict is 1)**an opposition or a tug-of-war between contradictory impulses .** according to colman "a conflict is 2)**the anticipated frustration entailed in the choice of either alternative**".
Query	conflict definition psychology
Ground-truth	the anticipated frustration entailed in the choice of either alternative.
PLM	conflict definition psychology.
Adapter	conflict is an opposition or a tug-of-war between contradictory impulses.

→ General meaning of *conflict*

- Added adapter incorrectly outputs *not the psychological meaning* but the *general meaning* of conflict due to overfitting to answering the general meaning in this few-shot setting.

Case study - QA

Passage	three types of conflicts are : 1. intrapersonal conflicts , 2. interpersonal conflicts and 3. unconscious conflicts . the word conflict has been derived from a latin word "conflicts" which means "strike two things at the same time" . conflict is [1)]**an opposition or a tug-of-war between contradictory impulses .** according to colman "a conflict is [2)]**the anticipated frustration entailed in the choice of either alternative**".
Query	conflict definition psychology
Ground-truth	the anticipated frustration entailed in the choice of either alternative.
PLM	conflict definition psychology.
Adapter	conflict is an opposition or a tug-of-war between contradictory impulses.
PLM with Condition	the meaning of conflict is (provided condition) the anticipated frustration entailed in the choice of either alternative.

- PLM generates the correct answer if the proper conditioning text is provided.
- The use of the added generator alone could ignore the PLM's knowledge.

Case study - QA

Passage	three types of conflicts are : 1. intrapersonal conflicts , 2. interpersonal conflicts and 3. unconscious conflicts . the word conflict has been derived from a latin word "conflicts" which means "strike two things at the same time" . conflict is [1)]**an opposition or a tug-of-war between contradictory impulses .** according to colman "a conflict is [2)]**the anticipated frustration entailed in the choice of either alternative**".
Query	conflict definition psychology
Ground-truth	the anticipated frustration entailed in the choice of either alternative.
PLM	conflict definition psychology.
Adapter	conflict is an opposition or a tug-of-war between contradictory impulses.
PLM with Condition	the meaning of conflict is (provided condition) the anticipated frustration entailed in the choice of either alternative.
Proposed STG	conflict is the anticipated frustration entailed in the choice of either alternative.

- Our proposed algorithm can generate the correct answer by explicitly leveraging both the PLM and the Adapter.

Selective Token Generation (STG)

Idea: ***Selectively generation*** *of each token either from the* ***task-specific adapter*** π_a *or the* ***PLM*** π_{PLM}

$\underbrace{\text{conflict is}}_{\pi_a}$ $\underbrace{\text{the anticipated frustration entailed in the choice of either alternative.}}_{\pi_{PLM}}$

Selective Token Generation (STG)

Idea: ***Selectively*** *(by using π_s)* ***generation*** *either from the* ***task-specific adapter*** π_a *or the* ***PLM*** π_{PLM}

$$\underbrace{\text{conflict is}}_{\substack{\pi_a \\ \uparrow \\ \pi_s}}\ \underbrace{\text{the anticipated frustration entailed in the choice of either alternative.}}_{\substack{\pi_{PLM} \\ \uparrow \\ \pi_s}}$$

Text generation as a RL problem

MDP

- State: $s_t = y_{1:t-1}$ (generated tokens so far)
- Action: t_{th} text token = $\mathrm{y_t} = a_t \in |\mathcal{V}|$, (e.g. $|\mathcal{V}| \approx 52K$ for GPT2)
- Reward: $r_t = r(s_t, a_t) = r(y_{1:t})$
  - $r_t = 0$, where $\mathrm{t} < \mathrm{T}$ and T is the sequence length
- Policy (generator & transition operator): $\pi_\theta(a_t|s_t)$

Text generation as a RL problem

MDP

- State: $s_t = y_{1:t-1}$ (generated tokens so far)
- Action: t_{th} text token = $\mathrm{y_t} = a_t \in |\mathcal{V}|$, (e.g. $|\mathcal{V}| \approx 52K$ for GPT2)
- Reward: $r_t = r(s_t, a_t) = r(y_{1:t})$
  - $r_t = 0, \text{where t} < \text{T and T is the sequence length}$
- Policy (generator & transition operator): $\pi_\theta(a_t|s_t)$

- Policy π_θ is trained to maximize the expected sum of future discounted rewards,

  $\mathbb{E}_{\tau \sim \pi_\theta}\left[\sum_{t=0}^{T} \gamma^t r_t\right]$,

  where γ ∈ $[0, 1]$ is the discount factor, and $\tau = \{s_t, a_t, r_t\}_{t=0}^{T}$ is the trajectory created by the MDP.

Text generation as a RL problem

MDP

- State: $s_t = y_{1:t-1}$ (generated tokens so far)
- Action: t_{th} text token = $y_t = a_t \in |\mathcal{V}|$, (e.g. $|\mathcal{V}| \approx 52K$ for GPT2)
- Reward: $r_t = r(s_t, a_t) = r(y_{1:t})$
  - $r_t = 0$, where $t < T$ and T is the sequence length
- Policy (generator & transition operator): $\pi_\theta(a_t|s_t)$

- Policy π_θ is trained to maximize the expected sum of future discounted rewards,

  $$\mathbb{E}_{\tau \sim \pi_\theta}\left[\sum_{t=0}^{T} \gamma^t r_t\right],$$

  where $\gamma \in [0, 1]$ is the discount factor, and $\tau = \{s_t, a_t, r_t\}_{t=0}^{T}$ is the trajectory created by the MDP.

- *Policy gradient loss for θ:

  $$\mathcal{L} = -\sum_{t=0}^{T} A^{\pi_\theta}(s_t, a_t) log\pi_\theta(a_t|s_t),$$

  A^{π_θ} is the advantage function in actor-critic framework

(* see more details in our paper)

STG as a RL problem

Hierarchical policy

- Selector: $\pi_{\theta_s}(i_t|s_t)$
- PLM policy: $\pi_{LM}(a_t|s_t)$
- Task specific policy: $\pi_{\theta_a}(a_t|s_t)$
- Action
  - $i_t \sim \pi_{\theta_s}(i_t|s_t)$,
  - t_{th} text token = $y_t = \begin{cases} a_t \sim \pi_{LM}(a_t|s_t) & if\ i_t = 0, \\ a_t \sim \pi_{\theta_a}(a_t|s_t) & if\ i_t = 1. \end{cases}$
- Hierarchical policy: $\pi_{\theta_h}(a_t|s_t; \theta_s, LM, \theta_a)$

STG as a RL problem

Hierarchical policy

- Selector: $\pi_{\theta_s}(i_t|s_t)$
- PLM policy: $\pi_{LM}(a_t|s_t)$
- Task specific policy: $\pi_{\theta_a}(a_t|s_t)$
- Action
  - $i_t \sim \pi_{\theta_s}(i_t|s_t)$,
  - t_{th} text token = $y_t = \begin{cases} a_t \sim \pi_{LM}(a_t|s_t) & if\ i_t = 0, \\ a_t \sim \pi_{\theta_a}(a_t|s_t) & if\ i_t = 1. \end{cases}$
- Hierarchical policy: $\pi_{\theta_h}(a_t|s_t; \theta_s, LM, \theta_a)$

- Trajectory $\tau = \{s_t, i_t, a_t, r_t\}_{t=0}^{T}$

STG as a RL problem

Hierarchical policy

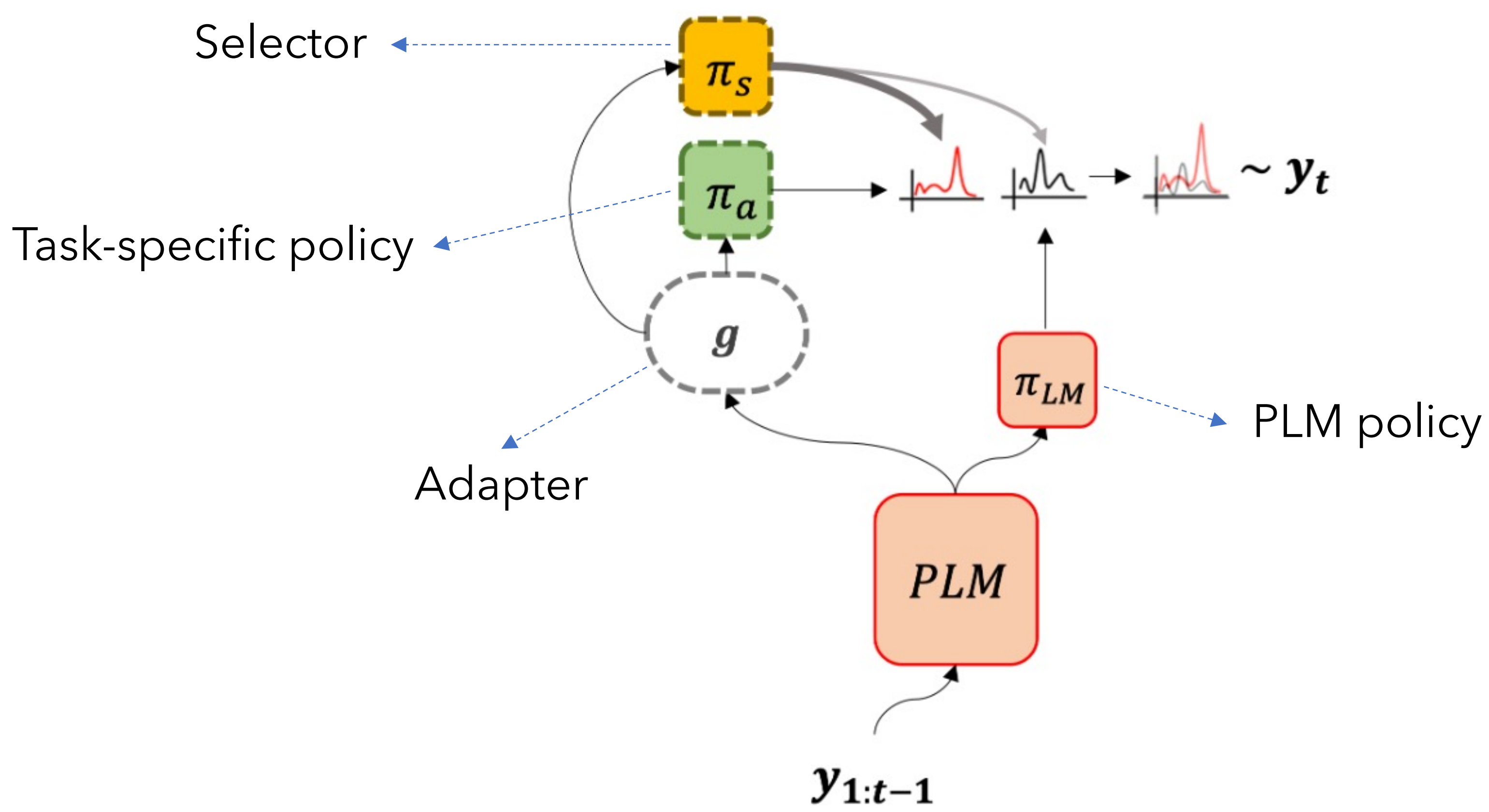
- Selector: $\pi_{\theta_s}(i_t|s_t)$
- PLM policy: $\pi_{LM}(a_t|s_t)$
- Task specific policy: $\pi_{\theta_a}(a_t|s_t)$
- Action
  - $i_t \sim \pi_{\theta_s}(i_t|s_t)$,
  - t_{th} text token = $y_t = \begin{cases} a_t \sim \pi_{LM}(a_t|s_t) & if\ i_t = 0, \\ a_t \sim \pi_{\theta_a}(a_t|s_t) & if\ i_t = 1. \end{cases}$
- Hierarchical policy: $\pi_{\theta_h}(a_t|s_t; \theta_s, LM, \theta_a)$
- Trajectory $\tau = \{s_t, i_t, a_t, r_t\}_{t=0}^{T}$
- *Policy gradient loss for θ_h :

$$\mathcal{L} = -\sum_{t=0}^{T} A^{\pi_{\theta_h}}\{\,\mathbb{I}_t[i_t = 0]\log(\pi_{\theta_s}(i_t|s_t)\mathrm{sg}[\pi_{LM}(a_t|s_t)]) + \mathbb{I}_t[i_t = 1]\log(\pi_{\theta_s}(i_t|s_t)\pi_{\theta_a}(a_t|s_t))\},$$

Where $\mathbb{I}_t[\cdot]$ is the indicator function, $A^{\pi_{\theta_h}}$ is the advantage function in actor-critic framework, and sg stands for the stop-gradient

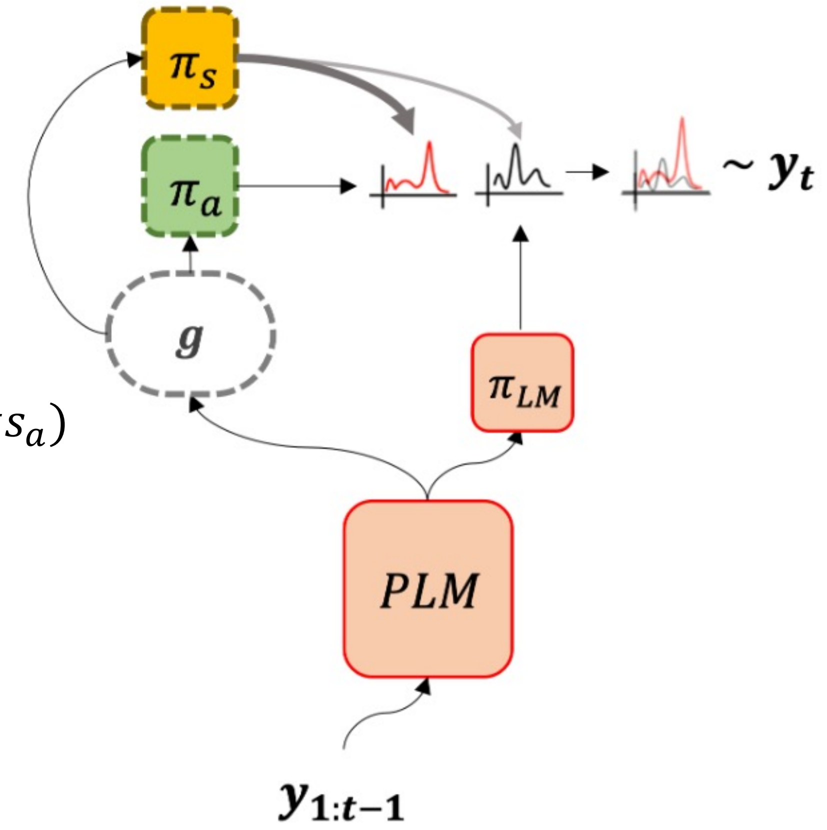
(*see more details in our paper)

Implementation of STG



Implementation of STG

- Adapter: $g\big(h_{LM};\theta_g\big) = \mathrm{LSTM}(h_{LM}(y_{<t}))$
- Selector: $\pi_s(i_t|h_{LM};\theta_s) = \mathrm{softmax}(m(g\big(h_{LM};\theta_g\big);\theta_s)$
- Task-specific policy: $\pi_a(\hat{y}_t|h_{LM};\theta_a) = \mathrm{softmax}(logits_{LM} + logits_a)$
    - $logits_a = W_a^{\mathrm{T}} g(h_{LM};\theta_g)$
    - $\theta_a \in \{W_a\},\ W_a \in R^{H\times|\mathcal{V}|}$
    - Auxiliary training [Zeldes et al., 2020]
        - ensure to learn from good initial policy
- Learnable parameters: $\{\theta_s, \theta_a, \theta_g\}$



Zeldes et al. 2020. Technical report: Auxiliary tuning and its application to conditional text generation

Experiments

- We evaluate STG on three different NLG tasks
    1. Data-to-Text
    2. Question Answering
    3. Text Summarization

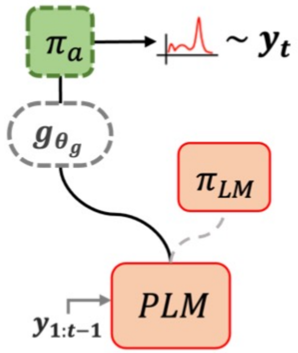
Baseline

1. Pretrained language model (PLM)
   - Fine-tuned GPT2-medium with MLE on few-shot data

   → Used as the PLM of models for covering large domain shift

2. Non-Selective token generation (Non-STG)
   - Non-STG-MLE
   - Non-STG-RL

Non-STG



3. Naïve Ensemble with PLM and Non-STG
   - NE(max): $\pi_{max} = Max(\pi_a, \pi_{LM})$
   - NE(mix): $\pi_{mix} = \frac{(\pi_a + \pi_{LM})}{2}$

   → A special case of STG

   - completely random selector
   - added generator trained independently for PLM

RL setup

- Actor-Critic framework
- Reward function
  - (Delexicalized) BLEU for Data-to-Text [Peng et al. 2020]
  - Averaged value of BLEU and ROUGE-L for Question Answering
  - ROUGE-L for Summarization [Paulus et al. 2017]

Peng et al. 2020. Few-shot natural language generation for task-oriented dialog
Paulus et al. 2017. A deep reinforced model for abstractive summarization

Data-to-Text

- FewShotWOZ [*Peng et al., 2020*]
  - A task that transforms structured data such as graphs or tables into natural language
  - Four available topics both for training & evaluation
    - Restaurant, Hotel, TV, Laptop
  - 50 training instances for each topic
  - 129, 78, 1379, and 680 testing instances for Restaurant, Hotel, Laptop, and TV, respectively

Peng et al. 2020. Few-shot natural language generation for task-oriented dialog

Data-to-Text

- FewShotWOZ [*Peng et al., 2020*]
  - A task that transforms structured data such as graphs or tables into natural language
  - Four available topics both for training & evaluation
    - Restaurant, Hotel, TV, Laptop
  - 50 training instances for each topic
  - 129, 78, 1379, and 680 testing instances for Restaurant, Hotel, Laptop, and TV, respectively

	Restaurant		Hotel		TV		Laptop	
Model	BLEU ↑	ERR ↓	BLEU ↑	ERR ↓	BLEU ↑	ERR ↓	BLEU ↑	ERR ↓
PLM	19.42	12.57	35.84	13.74	29.0	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	15.2	11.68	32.68	16.48	28.91	9.24	28.66	9.51
NE(mix)-RL	24.1	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**

Table 2: Data-to-Text performance on FewShotWOZ dataset.

Peng et al. 2020. Few-shot natural language generation for task-oriented dialog

Question Answering

- MS-MARCO [*Nguyen et al., 2016*]
  - A passage and a query are given, model should generate an answer with respect to the query by referring to the passage
  - We randomly sample various sizes of (50, 100, 500, 1,000 ≈ 1%, and 2,000) subset data from the train dataset over three different random seeds
  - Validation set (500) and test set (12,000) are shared

Nguyen et al. 2016. Ms marco: A human generated machine reading comprehension dataset

Question Answering

- MS-MARCO [*Nguyen et al., 2016]*
  - A passage and a query are given, model should generate an answer with respect to the query by referring to the passage
  - We randomly sample various sizes of (50, 100, 500, 1,000 ≈ 1%, and 2,000) subset data from the training dataset over three different random seeds
  - Validation set (500) and test set (12,000) are shared

	50 shot		100 shot		500 shot		1,000 shot		2,000 shot	
Model	BLEU	R-L	BLEU	R-L	BLEU	R-L	BLEU	R-L	BLEU	R-L
PLM	19.99	29.01	34.93	41.27	35.64	43.10	41.49	49.76	47.72	56.02
Non-STG-MLE	27.46	35.08	34.08	40.93	34.53	43.08	41.02	50.14	47.85	56.81
Non-STG-RL	20.07	28.94	35.08	41.28	35.08	42.78	41.25	49.97	48.00	56.83
NE(max)-MLE	27.21	34.95	34.76	41.87	34.69	43.93	41.11	50.77	47.65	57.22
NE(mix)-MLE	26.97	35.1	35.31	41.82	36.26	44.43	42.26	**51.14**	**48.44**	**57.3**
NE(max)-RL	20.05	28.9	35.0	41.16	35.14	42.94	41.51	50.54	47.58	57.06
NE(mix)-RL	20.69	29.62	35.11	41.33	35.93	43.52	42.29	50.84	48.28	57.02
STG	**33.33**	**39.59**	**36.3**	**43.24**	**37.37**	**44.53**	**42.76**	**51.19**	**48.42**	**57.3**

Table 3: *Averaged performances* for Question Answering on various few-shot subset data of MS-MARCO.

Nguyen et al. 2016. Ms marco: A human generated machine reading comprehension dataset

Question Answering

- MS-MARCO *[Nguyen et al., 2016]*
  - A passage and a query are given, model should generate an answer with respect to the query by referring to the passage
  - We randomly sample various sizes of (50, 100, 500, 1,000 ≈ 1%, and 2,000) subset data from the train dataset over three different random seeds for each size
  - Validation set (500) and test set (12,000) are shared

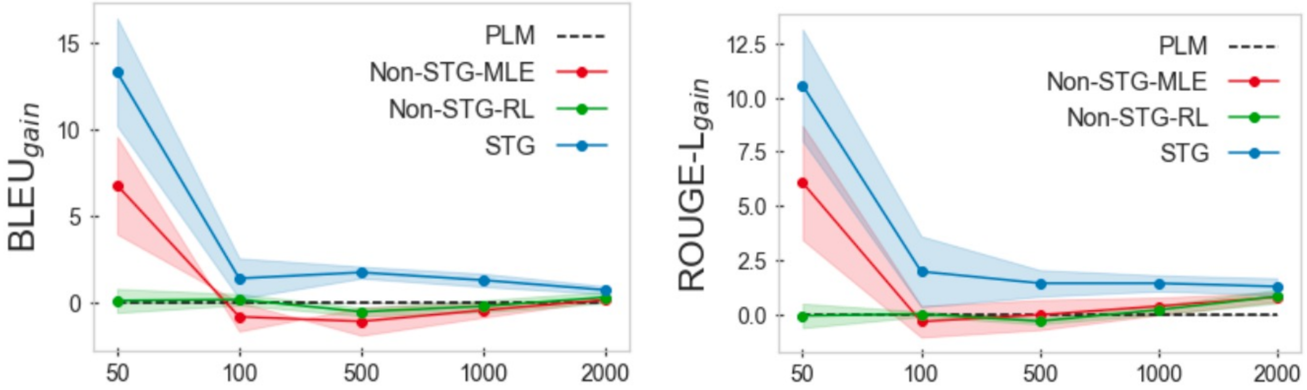


Figure 3: *Averaged performance gains* against the PLM for Question Answering on various few-shot subset data of MS-MARCO. The x-axis represents the size of the subset data and the shaded area represents a range of standard deviation over 3 randomly sampled subset data with different random seeds. STG provides significantly larger gains compared to Non-STGs on BLEU (Left) and ROUGE-L (Right).

Nguyen et al. 2016. Ms marco: A human generated machine reading comprehension dataset

Summarization

- CNN/DM [*See et al., 2017]*
  - Abstractive summarization task for long text generation
  - We randomly sample various sizes of (50, 100, 300, 1,500, and 3,000 ≈ 1%) subset data over three different random seeds for each size
  - Validation set (500) and test set are shared

See et al. 2017. Get to the point: Summarization with pointer-generator networks

Summarization

- CNN/DM *[See et al., 2017]*
  - Abstractive summarization task for long text generation
  - We randomly sample various sizes of (50, 100, 300, 1,500, and 3,000 ≈ 1%) subset data over three different random seeds for each size
  - Validation set (500) and test set are shared

	50 shot			100 shot			300 shot			1,500 shot			3,000 shot		
Model	R1	R2	R-L	R1	R2	R-L	R1	R2	R-L	R1	R2	R-L	R1	R2	R-L
PLM	14.67	4.57	10.69	16.58	5.28	12.05	19.38	7.08	13.74	30.19	11.27	21.21	33.05	12.96	23.39
Non-STG-MLE	15.39	4.81	11.09	17.09	5.41	12.3	18.9	6.87	13.36	30.34	11.32	21.2	33.19	12.98	23.39
Non-STG-RL	15.22	4.76	11.08	16.55	5.25	12.0	19.61	7.11	13.83	30.35	11.34	21.22	33.22	12.99	23.4
NE(max)-MLE	15.52	4.89	11.24	16.98	5.43	12.26	19.19	7.0	13.56	30.33	11.31	21.2	33.19	12.99	23.4
NE(mix)-MLE	15.4	4.83	11.16	16.88	5.4	12.22	19.45	7.07	13.75	30.32	11.31	21.23	33.11	12.99	23.41
NE(max)-RL	15.14	4.73	11.02	16.52	5.27	11.99	19.47	7.1	13.76	30.37	11.35	21.26	33.21	12.99	23.41
NE(mix)-RL	14.95	4.67	10.89	16.6	5.29	12.04	19.58	7.14	13.84	30.28	11.3	21.22	33.14	13.0	23.42
STG	**17.4**	**5.33**	**12.42**	**17.96**	**5.73**	**12.94**	**23.27**	**8.32**	**16.29**	**30.47**	**11.37**	**21.36**	**33.45**	**13.14**	**23.66**

Table 4: *Averaged performances* for Text Summarization on various few-shot subset data of CNN/DM.

See et al. 2017. Get to the point: Summarization with pointer-generator networks

Summarization

- CNN/DM *[See et al., 2017]*
  - Abstractive summarization task for long text generation
  - We randomly sample various sizes of (50, 100, 300, 1,500, and 3,000 ≈ 1%) subset data over three different random seeds for each size
  - Validation set (500) and test set are shared

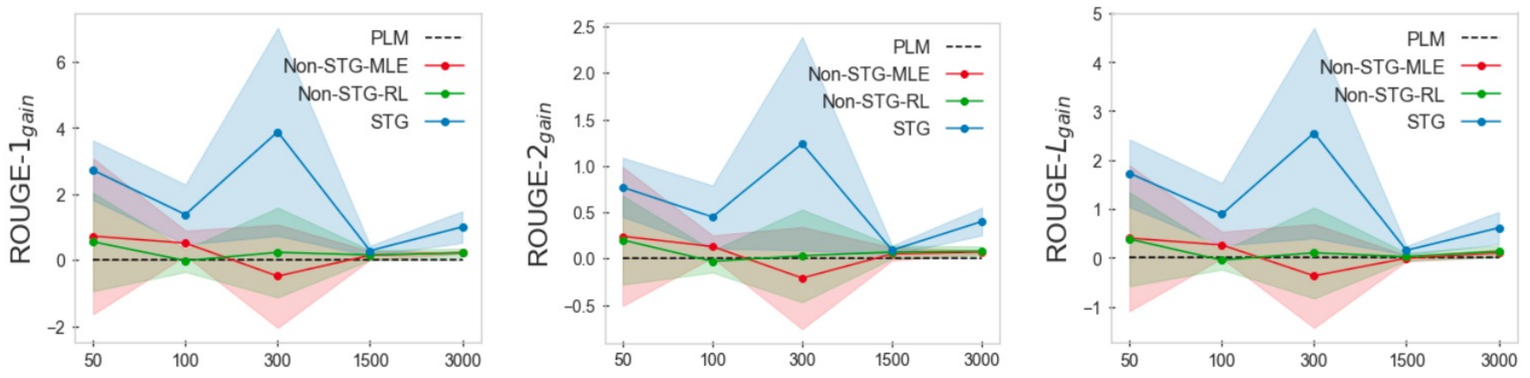


Figure 4: *Averaged performance gains* against the PLM for Text Summarization on various few-shot subset data of CNN/DM. The x-axis represents the size of the subset data and the shaded area represents a range of standard deviation over 3 randomly sampled subset data with different random seeds. STG provides significantly larger gains compared to Non-STGs on ROUGE-1 (Left), ROUGE-1 (Middle), and ROUGE-L (Right).

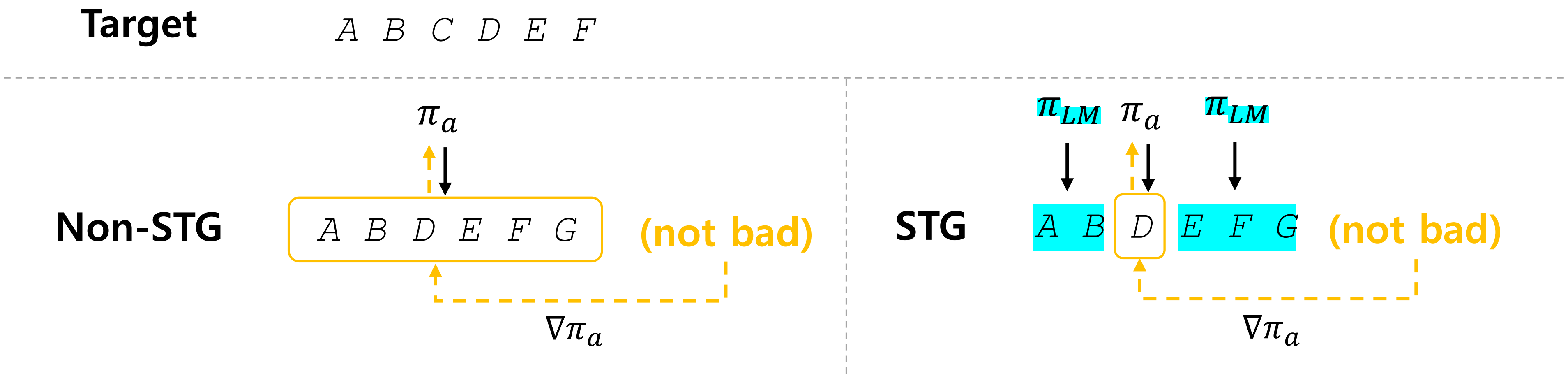
See et al. 2017. Get to the point: Summarization with pointer-generator networks

Advantages of STG

1. STG makes use of PLM **not only at the feature level but also the output distribution level** in text generation.
   - The PLM produces **task-general parts** while the adapter generates only **task-specific parts**.
   - It is beneficial in retaining strong linguistics and world knowledge of PLM

2. STG's search space is approximately decreased from $|\mathcal{V}|^{T}$ to $|\mathcal{V}|^{T-\bar{T}_{PLM}}$ where $\bar{T}_{PLM}$ is the average length of sequences generated by PLM.

3. STG is efficient in credit assignment.



Limitations & Future work

- Adapter
  - Relatively naive adapter which utilizes only top layer of PLM is used and this may lead to limited improvements as shown in the experiment of summarization.
  - Future work will consider more efficient adapters for covering a large domain shift and scaling.

- Efficient exploration
  - The fundamental limitation in STG is a high dependency on PLM
    - STG may nothing more than PLM when sufficient powerful PLM is used.
    - STG may nothing more than Non-STG when extremely poor PLM is used.

    → ***Balanced selection between PLM and adapter is required during in exploration in RL***
  - RL objective requires more training time than MLE objective (e.g. Prefix-Tuning) due to the auto-regressive sequence sampling during training.

    → ***Analysis on efficient exploration of STG is important for future works***

Conclusion

- A novel selective token generation between the PLM and the task-specific adapter is proposed for few-shot NLG.

- RL is applied to train both the policy selector and the task-specific adapter that is complementary to the PLM in text generation.

- Experimental results on various tasks of few-shot text generation show that the proposed method consistently and significantly improves the performances.

Thanks for your attention!