Neural Mask Generator: Learning to Generate Adaptive Word Maskings for Language Model Adaptation

Minki Kang[1*] **Moonsu Han**[1*] **Sung Ju Hwang**[1,2]

KAIST[1], Daejeon, South Korea
AITRICS[2], Seoul, South Korea
{zzxc1133, mshan92, sjhwang82}@kaist.ac.kr

Abstract

We propose a method to automatically generate a domain- and task-adaptive maskings of the given text for self-supervised pre-training, such that we can effectively adapt the language model to a particular target task (e.g. question answering). Specifically, we present a novel reinforcement learning-based framework which learns the masking policy, such that using the generated masks for further pre-training of the target language model helps improve task performance on unseen texts. We use off-policy actor-critic with entropy regularization and experience replay for reinforcement learning, and propose a Transformer-based policy network that can consider the relative importance of words in a given text. We validate our *Neural Mask Generator (NMG)* on several question answering and text classification datasets using BERT and DistilBERT as the language models, on which it outperforms rule-based masking strategies, by automatically learning optimal adaptive maskings. [1]

1 Introduction

The recent success of the *language model pre-training* approaches (Devlin et al., 2019; Peters et al., 2018; Radford et al., 2019; Raffel et al., 2019; Yang et al., 2019), which train language models on diverse text corpora with self-supervised or multi-task learning, have brought up huge performance improvements on several natural language understanding (NLU) tasks (Wang et al., 2019; Rajpurkar et al., 2016). The key to this success is their ability to learn generalizable text embeddings that achieve near optimal performance on diverse tasks with only a few additional steps of fine-tuning on each downstream task.

Most of the existing works on language model aim to obtain a universal language model that can address nearly the entire set of available natural language tasks on heterogeneous domains. Although this train-once and use-anywhere approach has been shown to be helpful for various natural language tasks (Devlin et al., 2019; Radford et al., 2019; Dong et al., 2019; Raffel et al., 2019), there have been considerable needs on adapting the learned language models to domain-specific corpora (e.g. healthcare or legal). Such domains may contain new entities that are not included in the common text corpora, and may contain only a small amount of labeled data as obtaining annotation on them may require expert knowledge. Some recent works (Sun et al., 2019a; Lee et al., 2019; Beltagy et al., 2019; Gururangan et al., 2020) suggest to further pre-train the language model with self-supervised tasks on the domain-specific text corpus for adaptation, and show that it yields improved performance on tasks from the target domain.

Masked Language Models (MLMs) objective in BERT (Devlin et al., 2019) has shown to be effective for the language model to learn the knowledge of the language in a bi-directional manner (Vaswani et al., 2017). In general, masks in MLMs are sampled at random (Devlin et al., 2019; Liu et al., 2019c), which seems reasonable for learning a generic language model pre-trained from scratch, since it needs to learn about as many words in the vocabulary as possible in diverse contexts.

However, in the case of further pre-training of the already pre-trained language model, such a conventional selection method may lead a domain adaptation in an inefficient way, since not all words will be equally important for the target task. Repeatedly learning for uninformative instances thus will be wasteful. Instead, as done with instance selection (Ngiam et al., 2018; Jiang et al., 2018; Yoon et al., 2019; Zhu et al., 2019), it will be more effective if the masks focus on the most important words for the target domain, and for the specific

[*] Equal contribution.
[1] Code is available at github.com/Nardien/NMG.

Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, pages 6102–6120, November 16–20, 2020. ©2020 Association for Computational Linguistics

NLU task at hands. How can we then *obtain* such a masking strategy to train the MLMs?

Several works (Joshi et al., 2019; Sun et al., 2019b,c; Glass et al., 2019) propose rule-based masking strategies which work better than random masking (Devlin et al., 2019) when applied to language model pre-training from scratch. Based on those works, we assume that adaptation of the pre-trained language model can be improved via a *learned* masking policy which selects the words to mask. Yet, existing models are inevitably suboptimal since they do not consider the target domain and the task. To overcome this limitation, in this work, we propose to adaptively generate mask by learning the optimal masking policy for the given task, for the task-adaptive pre-training (Gururangan et al., 2020) of the language model.

As described in Figure 1, we want to further pre-train the language model on a specific task with a task-dependent masking policy, such that it directs the solution to the set of parameters that can better adapt to the target domain, while task-agnostic random policy leads the model to an arbitrary solution. To tackle this problem, we pose the given learning problem as a meta-learning problem where we learn the task-adaptive mask-generating policy, such that the model learned with the masking strategy obtains high accuracy on the target task. We refer to this meta-learner as the **N**eural **M**ask **G**enerator (NMG). Specifically, we formulate mask learning as a bi-level problem where we pre-train and fine-tune a target language model in the inner loop, and learn the NMG at the outer loop, and solve it using renforcement learning. We validate our method on diverse NLU tasks, including question answering and text classification. The results show that the models trained using our NMG outperforms the models pre-trained using rule-based masking strategies, as well as finds a proper adaptive masking strategy for each domain and task.

Our contribution is threefold:

- We propose to learn the mask generating policy for further pre-training of masked language models, to obtain optimal maskings that focus on the most important words for the given text domain and the NLU task.

- We formulate the problem of learning the task-adaptive mask generating policy as a bi-level meta-learning framework which learns the LM in the inner loop, and the mask generator at the outer loop using reinforcement learning.

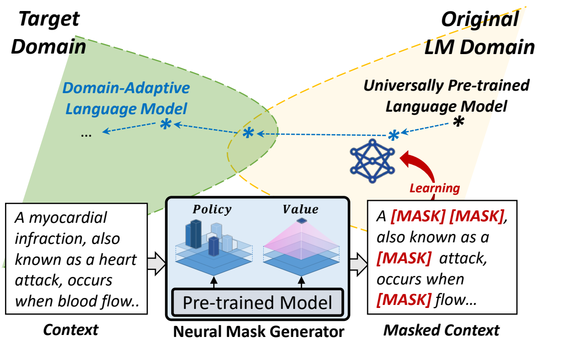

Figure 1: **Concept.** Pre-training on domain text leads the language model parameters to adapt to the given target domain. We assume that adjusting the masking policy of the MLM objective affects the training trajectory of the language model, such that it moves towards a better solution space for the target domain. This illustration of the solution spaces for the two domains is motivated by (Gururangan et al., 2020).

- We validate our mask generator on diverse tasks across various domains, and show that it outperforms heuristic masking strategies by learning an optimal task-adaptive masking for each LM and domain. We also perform empirical studies on various heuristic masking strategies on the language model adaptation.

2 Related Work

Language Model Pre-training Ever since Howard and Ruder (2018) suggested language model pre-training with multi-task learning, inspired by the success of fine-tuning on ImageNet pre-trained models on computer vision tasks (Liu et al., 2019b), research on the representation learning for natural language understanding tasks have focused on obtaining a global language model that can generalize to any NLU tasks. A popular approach is to use self-supervised pre-training tasks for learning the contextualized embedding from large unannotated text corpora using auto-regressive (Peters et al., 2018; Yang et al., 2019) or auto-encoding (Devlin et al., 2019; Liu et al., 2019c) language modeling. Following the success of the Masked Language Model (MLM) from (Devlin et al., 2019; Liu et al., 2019c), several works have proposed different model architecture (Lan et al., 2019; Raffel et al., 2019; Clark et al., 2020) and pre-training objectives (Sun et al., 2019b,c; Liu et al., 2019c; Joshi et al., 2019; Glass et al., 2019; Dong et al., 2019), to improve upon its performance. Some works have also proposed alternative masking policies for the MLM pre-training over random sampling, such as SpanBERT (Joshi et al., 2019) and ERNIE (Sun et al., 2019b). Yet, none of the existing approaches have tried to *learn* the task-adaptive mask in a context-dependent manner which is the problem we target in this work.

Language Model Adaptation Pre-training the language model on the target domain, then fine-tuning on downstream tasks, is the most simple yet successful approach for adapting the language model to a specific task. Some studies (Lee et al., 2019; Beltagy et al., 2019; Whang et al., 2019) have shown the advantage of further pre-training the language model on a large unlabeled text corpus collected from a specific domain. Moreover, Sun et al. (2019a) and Han and Eisenstein (2019) investigate the effectiveness of further pre-training of the language model on small domain-related text corpora. Recently, Gururangan et al. (2020) integrates prior works and defines domain-adaptive pre-training and task-adaptive pre-training, showing that domain adaptation of the language model can be done with additional pre-training with the MLM objective on a domain-related text corpus, as well as a smaller but directly task-relevant text corpus.

Meta-Learning Meta-learning (Thrun and Pratt, 1998) aims to train the model to generalize over a distribution of tasks, such that it can generalize to an unseen task. There exist large number of different approaches to train the meta-learner (Santoro et al., 2016; Vinyals et al., 2016; Ravi and Larochelle, 2017; Finn et al., 2017; Liu et al., 2019a). However, existing meta-learning approaches do not scale well to the training of large models such as masked language models. Thus, instead of the existing meta-learning method such a gradient based approach, we formulate the problem as a bi-level problem (Franceschi et al., 2018) of learning the language model in the inner loop and the mask at the outer loop, and solve it using reinforcement learning. Such optimization of the outer objective using RL is similar to the formulation used in previous works on neural architecture search (Zoph and Le, 2017; Zoph et al., 2018).

3 Problem Statement

We now describe how to formulate the problem of *learning to generate masks* for the Masked Language Model (MLM) as a bi-level optimization problem. Then, we describe how we can reformulate it as a reinforcement learning problem.

3.1 Masked Language Model

For pre-training of the language models, we need an unannotated text corpus $\boldsymbol{S} = \{s^{(1)}, \cdots, s^{(T)}\}$. Here the $s = [w_1, w_2, \cdots, w_N]$ is the context, whose element $\boldsymbol{w_i} \in s$ is a single word or token. To formulate the meta-learning objective, we assume each context corpus as the part of the task $\mathcal{T} = \{\boldsymbol{S}, D_{tr}, D_{te}\}$ consisting of the context and its corresponding task dataset. For the MLM, we need to generate a noisy version of s, which we denote as $\hat{s}$. Let z_i be the indicator of masking i-th word of s. If $z_i = 1$, i-th word is replaced with the [MASK] token. The objective of the MLM then is to predict the original token of each [MASK] token. Therefore, we can formulate this problem as follows:

$$\min_{\theta} \mathcal{L}_{MLM}(\theta; \boldsymbol{S}, \hat{\mathbf{S}}) = \sum_{s, \hat{s} \in \boldsymbol{S}, \hat{\mathbf{S}}} -\log p_\theta(\bar{s}|\hat{s})$$

where θ is the parameter of the language model and $\bar{s}$ is the original tokens of each [MASK] token, and $\hat{\mathbf{S}}$ is the set of masked contexts. Following the formulation from (Yang et al., 2019), we can approximate the MLM objective with z_i as follows:

$$\max_{\theta} \log p_\theta(\bar{s}|\hat{s}) \approx \sum_{i=1}^{N} z_i \log p_\theta(w_i|\hat{s}) \quad (1)$$

$$p_\theta(w_i|\hat{s}) = \frac{\exp(H_\theta(\hat{s})_i^\top e(w_i))}{\sum_{w'} \exp(H_\theta(\hat{s})_i^\top e(w'))}$$

where $H_\theta(\hat{s})$ indicates the contextualized representation of $\hat{s}$ from the language model layer (e.g. Pretrained Transformer layers in (Devlin et al., 2019)), and $e(w_i)$ denotes the embedding of word w_i from the last prediction layer. Our objective then is to learn the optimal policy for determining each mask indicator z_i, which we will describe in detail in the next subsection.

3.2 Bi-level formulation

We now describe how to formulate this learning problem as a bi-level problem consisting of inner and outer level objectives. Consider that z_i can be represented using an arbitrary function $\mathcal{F}$ parameterized with λ:

$$\pi_\lambda(a_t = i|s) = \mathcal{F}_\lambda(w_i) \quad (2)$$

$$\mathbf{a}_\lambda = \arg\max_{a} \prod_{t} \pi_\lambda(a_t|s) \quad (3)$$

$$z_{\lambda,i} = \begin{cases} 1, & \text{if } i \in \mathbf{a}_\lambda \\ 0, & \text{if } i \notin \mathbf{a}_\lambda \end{cases} \quad (4)$$

where $\pi_\lambda(a_t = i|s)$ is the probability of masking i-th word w_i from s, and a_λ indicates the list of word indices to be masked and $z_{\lambda,i}$ is the mask indicator parameterized by the parameter λ. The details of corresponding equations will be described in section 3.3. Therefore, the MLM objective has been slightly changed from its original form, into the following objective:

$$\begin{aligned}\theta(\lambda) &= \arg\min_{\theta} \mathcal{L}_{MLM}(\theta, \lambda; \boldsymbol{S}, \hat{\mathbf{S}}) \\ &= \arg\max_{\theta} \sum_{s \in \boldsymbol{S}} \log p_\theta(\bar{s}|\hat{s}) \\ &\approx \arg\max_{\theta} \sum_{s \in \boldsymbol{S}} \sum_{i=1}^{N} z_{\lambda,i} \log p_\theta(w_i|\hat{s})\end{aligned} \tag{5}$$

where the masked context $\hat{s}$ is parameterized by the parameter λ. Now assume that we have found the optimal parameter $\theta(\lambda)$ for language model from equation 5. Then, we need to fine-tune the language model on the downstream task. Although linear heads used in both pre-training and fine-tuning are different, we describe the parameter of both models as $\theta(\lambda)$ for simplicity. Following the bi-level framework notation described in (Franceschi et al., 2018), the inner objective function for fine-tuning can be written as follows:

$$\min_{\theta(\lambda)} \mathcal{L}_{train}(\theta(\lambda), \lambda) = \sum_{(x,y) \in D_{tr}} l(g_{\theta(\lambda)}(x), y) \tag{6}$$

where D_{tr} is a training dataset, l is the loss function of the supervised learning, and $g_{\theta(\lambda)}$ is the function representation of downstream task solver model. In case of question answering task, each x consists of a context and corresponding question and y is the corresponding answer spans.

Assume that we find optimal parameter $\theta^*(\lambda)$ from supervised task fine-tuning. Then, the final outer-level objective can be described as follows:

$$\begin{aligned}\min_{\lambda} \mathcal{L}_{test}(\lambda, \theta^*(\lambda)) &= \sum_{(x,y) \in D_{te}} l(g_{\theta^*(\lambda)}(x), y) \\ s.t. \quad \theta^*(\lambda) &= \arg\min_{\theta(\lambda)} \mathcal{L}_{train}(\theta(\lambda), \lambda) \\ and \quad \theta(\lambda) &= \arg\min_{\theta} \mathcal{L}_{MLM}(\theta, \lambda)\end{aligned} \tag{7}$$

This will allow us to obtain the optimal parameter λ^* which minimizes the task objective function on a test dataset D_{te}.

Although the outer objective is differentiable, we formulate the optimization problem of the outer objective as a reinforcement learning problem to avoid excessive computation cost caused by the two constraint terms.

Justification of Reinforcement Learning In this paragraph, we explain why we use the Reinforcement Learning (RL) instead of the differentiable method to train the parameter λ. As indicated in the equation 7, our inner loop includes consecutive two steps of language model training. The NMG model λ is addressed to the pre-training step rather than the task fine-tuning step. Therefore, the direct differentiation of the outer objective contains two second-derivative terms for both the MLM loss $\mathcal{L}_{MLM}$ and the task train loss $\mathcal{L}_{train}$. With a single-step approximation, the derivative of the outer objective is approximated as follows:

$$\begin{aligned}&\nabla_\lambda \mathcal{L}_{test}(\lambda, \theta^*(\lambda)) \\ &\approx \nabla_\lambda \mathcal{L}_{test}(\lambda, \tilde{\theta}^*(\lambda)) \\ &- \alpha \nabla^2_{\lambda,\theta} \mathcal{L}_{MLM}(\theta, \lambda) \nabla_\theta \mathcal{L}_{test}(\tilde{\theta}^*(\lambda), \lambda) \\ &- \beta \nabla^2_{\lambda,\tilde{\theta}(\lambda)} \mathcal{L}_{train}(\tilde{\theta}(\lambda), \lambda) \nabla_{\tilde{\theta}(\lambda)} \mathcal{L}_{test}(\tilde{\theta}(\lambda), \lambda)\end{aligned}$$

where $\tilde{\theta}(\lambda) = \theta - \alpha \nabla_\theta \mathcal{L}_{MLM}(\theta, \lambda), \tilde{\theta}^*(\lambda) = \tilde{\theta}(\lambda) - \beta \nabla_{\tilde{\theta}(\lambda)} \mathcal{L}_{train}(\tilde{\theta}(\lambda), \lambda)$ are approximated parameters, and α, β are learning rates. Such gradient estimation requires high computational costs since it includes the computation of Hessian-product vectors of the massive language model's parameters approximated as 110 millions (Finn et al., 2017; Devlin et al., 2019).

Instead, we can address the first-order approximation ($\alpha = 0, \beta = 0$) to the derivative of the outer objective to avoid second-order derivative computation as follows:

$$\nabla_\lambda \mathcal{L}_{test}(\lambda, \theta^*(\lambda)) \approx \nabla_\lambda \mathcal{L}_{test}(\theta)$$

where $\theta^*(\lambda)$ is approximated to θ. Such approximation trivially results in a meaningless optimization since it ignores the pre-training step induced by the parameter λ, which decides the masking policy (Liu et al., 2019a). Therefore, we approach solving this optimization problem with RL instead of the differential method to avoid such an issue. In the next section, we introduce how we formulate this problem as the RL with the outer objective as a non-differentiable reward.

3.3 Reinforcement learning formulation

We now propose a reinforcement learning (RL) framework, which given the context s as the state, decides on the actions $\mathbf{a} = \{a_1, \cdots, a_T\}$ where T is the number of masked tokens in the given context, each $a_t \in [1, N]$ is the token index that indicates a decision on masking the token w_{a_t} following equation 4, and $a_1 \neq a_2 \neq \cdots \neq a_T$. The objective of the RL agent then is to find an optimal masking policy that minimizes $E(\lambda, \theta)$ from the section 3.2. In addition, minimizing $E(\lambda, \theta)$ on D_{te} can be seen as maximizing its performance. Therefore, the objective is the maximization of the performance of the model on D_{te}. We can induce it by setting the reward R as the accuracy improvement on D_{te}. We will describe the detail of the reward design in section 4.

In general RL formulation (Sutton and Barto, 2018) following Markov Decision Process (MDP), state transition probability can be described as $p(s_{t+1}|s_t, a_t)$. The probability of masking T tokens is formulized as $p(\hat{s}|s) = \prod_{t=1}^{T} p(s_{t+1}|s_t, a_t)$, where $\hat{s}$ consists of T number of [MASK] tokens. Although the representation of s_t and s_{t+1} are slightly different because of the addition of [MASK] token, we can approximate them as $s_t \approx s_{t+1}$ following the approximation in equation 1, which inductively approximates representations after each word masking as a representation of original context.

Therefore, we can approximate the probability of masking T tokens from the MDP problem to the problem without state-transition formulation as follows:

$$p(\hat{s}|s) = \prod_{t=1}^{T} p(s_{t+1}|s_t, a_t) \approx \prod_{t=1}^{T} \pi(a_t|s)$$

where a_t denotes the index of masked word included in the context s and $\pi(\cdot|s)$ is the masking policy of the agent. By this approximation, we do not need to consider the trajectory along the temporal horizon, which makes the problem much easier. Instead, we approximate the problem as the task of selecting multiple discrete actions simultaneously for the same state. We approximate the policy with neural network parameterized by λ as $\pi_\lambda(a|s)$. As in equations 3 and 4, the mask $z_{\lambda,i}$ is determined by actions generated from the neural policy $\pi_\lambda(a|s)$. In the next section, we will describe how to train the Neural Mask Generator to generate the neural policy $\pi_\lambda(a|s)$ maximizing the reward R using RL in Section 4.

4 Neural Mask Generator

In this section, we describe our model, *Neural Mask Generator (NMG)*, which learns the masking policy to generate an optimal mask on an unseen context using deep RL with the detailed descriptions of the framework setup. The overview of the meta mask generator framework is shown in Figure 2. For detailed descriptions of the approaches, the procedures of both training and test phase, and algorithm, please **see Appendix A**.

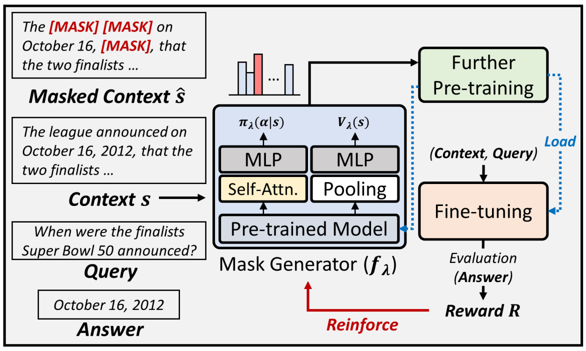

Figure 2: The overview of the meta-training framework for our Neural Mask Generator (NMG). In each episode, masked contexts by the NMG are used for further pre-training. Then, the further pre-trained language model is used in both the mask generator and fine-tuning.

4.1 Reinforcement Learning Details

Model Architecture The probability of selecting i-th word w_i of s as t-th action a_t can be described as $\pi_\lambda(a_t = i|s)$, where $\sum_i \pi_\lambda(a_t = i|s) = 1$. Instead of $\mathcal{F}_\lambda$ in equation 2, the neural policy from the NMG is given as follows:

$$\pi_\lambda(a_t = i|s) = \frac{\exp(f_\lambda(H_{\theta'}(s)_i))}{\sum_t \exp(f_\lambda(H_{\theta'}(s)_t))}$$

where f_λ is a deep neural network parameterized by λ, which has a self-attention layer (Vaswani et al., 2017) followed by linear layers with gelu activation (Hendrycks and Gimpel, 2016), and $H_{\theta'}(s)_i$ is the contextualized representation of w_i of context s from the frozen language model layer θ'. Note that θ' is shared with the target language model θ and not trained during the NMG training. Further, $f_\lambda(H_{\theta'}(s)_i)$ outputs the scalar logit of the w_i, and the final probabilistic policy is computed by the softmax function.

Training Objective We train the NMG model using the Advantage Actor-Critic method (Mnih et al., 2016) with the value estimator. Furthermore, we resort to off-policy learning (Degris et al., 2012) such that the agent can explore the optimal policy based on its past experiences. To this end, we leverage a prioritized experience replay, in which we store every state, action, reward, and old-policy pairs (Mnih et al., 2013), and sample them based on their absolute value of the advantage (Schaul et al., 2016). We use importance sampling to estimate the value of the target policy π_λ with the samples from the behavior policy $\pi_{\lambda_{old}}$, which are sampled from the replay buffer (Degris et al., 2012; Schulman et al., 2015, 2017; Wang et al., 2017). To sum up, the objectives for both policy and value network are as follows:

$$\mathcal{L}_{policy} = \sum_{(s,a,R,\pi_{\lambda_{old}})} - \frac{\pi_\lambda(a|s)}{\pi_{\lambda_{old}}(a|s)}(R - V_\lambda(s)) - \alpha\, \mathcal{H}(\pi_\lambda(a|s)) \tag{8}$$

$$\mathcal{L}_{value} = \sum_{(s,a,R,\pi_{\lambda_{old}})} \frac{1}{2}(R - V_\lambda(s))^2 \tag{9}$$

where the $(s, a, R, \pi_{\lambda_{old}})$ is set of sampled replays from the replay buffer, $\mathcal{H}$ is an entropy function $\mathcal{H}(\pi_\lambda(a|s)) = -\sum_t \pi_\lambda(a_t|s) \log \pi_\lambda(a_t|s)$, and α is a hyperparameter for entropy regularization. $V_\lambda(s)$ is an estimated value of the state s. The value network consists of linear layers with activation function after mean pooling $\frac{1}{N}\sum_t H_{\theta'}(s)_t$.

To summarize, the outer-level objective for updating the NMG parameter λ can be written as follows:

$$\min_\lambda \mathcal{L}_{policy} + \mathcal{L}_{value} \tag{10}$$

At each episode of meta-training, we update the NMG parameter λ by optimizing the above objective function.

Reward Design and Self-Play As in Section 3.3, the reward function is considered as the accuracy improvement on the test set D_{te}. Therefore, the pre-training step (equation 5) and the fine-tuning then evaluation step (equation 6, 7) should be done to get the reward in every episodes.

Since using the full size of dataset in the inner loop is generally not feasible, we randomly sample smaller sub-task $\mathcal{T}' = \{\boldsymbol{S}', D'_{tr}, D_{val}\}$ from $\mathcal{T}$ at every episode. For the evaluation, we randomly split the training set D_{tr} to generate a hold-out validation set D_{val} and replace D_{te} in equation 7 to D_{val} while meta-training where D_{te} is unobservable. We use a sufficiently large hold-out validation set D_{val} to prevent the masking policy from overfitting to D_{val}. We assume that the meta-learner NMG also performs well on $\mathcal{T}$ if it is trained on diverse sub-tasks $\mathcal{T}'$ where $|\mathcal{T}| \gg |\mathcal{T}'|$.

The problem to be considered for using diverse sub-tasks is that the NMG model encounters different sub-task $\mathcal{T}'$ at every new episode. Since $\mathcal{T}'$ determines the state distribution $\boldsymbol{S}'$ and the data D_{tr} to be trained, it results in the reward scale problem that the expectation of the validation accuracy on D_{val} varies depending on the composition of $\mathcal{T}'$ then makes it harder to evaluate the performance increment of the neural policy across episodes.

To address this problem, we introduce the random policy as an opponent policy to evaluate the neural policy relative to it. Therefore, the reward R is defined as $sgn(r - b)$, where sgn is the sign function, r and b are the accuracy score on D_{val} from neural and random policy respectively. However, the random policy may be too weak as the opponent. To overcome this limitation, we add another neural policy as an additional opponent to induce the zero-sum game of two learning agent by the concept of the self-play algorithm (Wei et al., 2017; Grau-Moya et al., 2018).

Then, three distinct policies are compared with each other during episodes and two neural policies are individually trained. Furthermore, in each neural agent training, only actions corresponding to disjoint comparing with others are stored in a global replay buffer for a more accurate reward assignment of each action.

Continual Adaptive Learning For fair comparison at every episode, the same language model should be used to evaluate the policy. Initializing the language model before the start of each inner loop can be the simplest choice to handle this. However, since we pre-train the language model for only few steps during each episode of meta-training, the model is always evaluated around the original language model domain (see Figure 1). To avoid this, at each episode, the language model which is pre-trained by the NMG model of former step is continually loaded instead of the fixed checkpoint except for the first episode. By this, we intend that our agent learns the optimal policy of various environments. Furthermore, the agent can learn dynamic policy based on the learned degree of the target language model.

5 Experiment

We now experimentally validate our Neural Mask Generator (NMG) model on multiple NLU tasks, including question answering and text classification tasks using two different language models, and analyze its behaviors. In Section 5.1, we evaluate the NMG with several baselines. Then, we evaluate the effect of the specific design choices made for our model through ablation studies in Section 5.2.

Finally, we analyze how the policy learned by the NMG model works in Section 5.3.

Tasks and Datasets For question answering, we use three datasets, namely SQuAD v1.1 (Rajpurkar et al., 2016), NewsQA (Trischler et al., 2017), and emrQA (Pampari et al., 2018) to validate our model. We use the MRQA[2] version for both SQuAD and NewsQA to sustain a coherency. We also preprocess emrQA to fit the format of other datasets. We use a standard evaluation metric named Exact-Matchs (EM) and F1 score for question answering task. For text classification, we use IMDb (Maas et al., 2011) and ChemProt (Kringelum et al., 2016), following the experimental settings of (Gururangan et al., 2020).

Baselines According to Devlin et al. (2019) and Joshi et al. (2019), training the language models with difficult objectives is much more beneficial when pre-training from scratch. To test whether it is also the case for task-adaptive pre-training, we experiment with two heuristic masking strategies, which we refer to as whole-random and span-random. In addition, we also tested the named entity masking proposed by Sun et al. (2019b) (entity-random). Below is the complete list of the heuristic baselines we compare against in the experiments.

1) No-PT A baseline without any further pre-training of the language model.

2) Random A random masking strategy introduced in BERT (Devlin et al., 2019).

3) Whole-Random A random masking strategy which masks the entire word instead of the token (sub-word). This method is introduced by the authors of BERT (Devlin et al., 2019)[3].

4) Span-Random A random masking strategy which selects multiple consecutive tokens.

5) Entity-Random A random masking strategy which selects named entities with highest priorities, then randomly selects other tokens.

6) Punctuation-Random A random masking strategy which selects punctuation tokens first, then randomly selects other tokens.

Implementation Details For the language model θ, we use the same hyperparameters and architecture with DistilBERT (Sanh et al., 2019) model (66M params) and BERT_{BASE} (Devlin et al., 2019) model (110M params). Our implementation is based on the huggingface's Pytorch implementation version (Wolf et al., 2019; Paszke et al., 2019). We load the pre-trained parameters from the checkpoint of each language model in meta-testing and the first episode of meta-training. As for the text corpus S to pre-train the language model, we use the collection of contexts from the given NLU task. We only use the Masked Language Model (MLM) objective for further pre-training. In the initial stage of meta-training, the NMG randomly selects actions for exploration. In meta-testing, it takes maximum probability indices as actions. We describe the details of language model training in **Appendix B** since we use different settings for each task and experiments. As for reinforcement learning (RL), we use the off-policy actor-critic method described in Section 4. For more details of the reinforcement learning framework, please **see Appendix B**.

5.1 Results

First of all, we need to discuss how the MLM works on the language model pre-training. In practice, the Cloze task (Taylor, 1953) benefits when words that need to learn are masked. However, for the MLM, we observed that the masking prevents learning the masked words in the language model. Rather the MLM learns representations and relations between non-masked words by predicting the masked words using them. In the case of adaptive pre-training, learning the domain-specific vocabulary is crucial for the domain adaptation. Therefore, masking out trivial words (e.g. Punctuations) may be more beneficial than masking out unique words (e.g. Named Entities) for the language model to learn the knowledge of a new domain. To see how it works, we experiment for punctuation-random, which masks only the punctuations, which are clearly useless for the domain adaptation.

In Table 1 and 2, we report the performance of baselines and our model on both question answering and text classification tasks. From the baseline results of Table 1, we speculate on the important aspects of a masking strategy for better adaptation. The whole-word, span and entity maskings often lead to better results since it makes the MLM objective more difficult and meaningful (Joshi et al., 2019; Sun et al., 2019b). For instance, in emrQA, most words are tokenized to sub-words since contexts include a lot of unique words such as medical terminologies. Therefore, whole-word masking could be most suitable for domain adaptation on

[2] https://mrqa.github.io/
[3] https://github.com/google-research/bert

Base LM	Model	SQuAD EM	SQuAD F1	emrQA EM	emrQA F1	NewsQA EM	NewsQA F1
BERT	No PT	$80.63_{0.32}$	$88.18_{0.25}$	$70.58_{0.09}$	$76.60_{0.29}$	$\underline{51.66}_{0.31}$	$\underline{66.23}_{0.16}$
	Random	$80.64_{0.02}$	$88.14_{0.10}$	$71.40_{1.01}$	$77.31_{0.78}$	$51.46_{0.19}$	$66.21_{0.22}$
	Whole	$80.73_{0.23}$	$88.20_{0.11}$	$\underline{71.65}_{0.33}$	$\mathbf{77.84}_{\mathbf{0.31}}$	$51.12_{0.28}$	$65.94_{0.51}$
	Span	$80.66_{0.19}$	$88.19_{0.11}$	$70.87_{0.38}$	$76.76_{0.33}$	$51.17_{0.63}$	$65.93_{0.50}$
	Entity	$\underline{80.79}_{0.13}$	$\underline{88.23}_{0.22}$	$71.50_{0.34}$	$\underline{77.57}_{0.27}$	$51.55_{0.43}$	$65.44_{0.34}$
	Punct.	$80.60_{0.24}$	$88.07_{0.25}$	$71.17_{0.58}$	$77.08_{0.57}$	$51.47_{0.37}$	$66.18_{0.36}$
	NMG	$\mathbf{80.83}_{\mathbf{0.20}}$	$\mathbf{88.28}_{\mathbf{0.21}}$	$\mathbf{71.84}_{\mathbf{0.68}}$	$77.49_{0.55}$	$\mathbf{51.81}_{\mathbf{0.33}}$	$\mathbf{66.57}_{\mathbf{0.48}}$
DistilBERT	No PT	$76.75_{0.41}$	$85.13_{0.26}$	$68.52_{0.39}$	$75.00_{0.53}$	$\underline{48.61}_{0.39}$	$\underline{63.45}_{0.56}$
	Random	$76.46_{0.35}$	$84.92_{0.17}$	$69.02_{0.40}$	$75.70_{0.29}$	$48.52_{0.46}$	$63.06_{0.21}$
	Whole	$76.48_{0.27}$	$84.96_{0.15}$	$\underline{69.64}_{0.39}$	$\underline{76.16}_{0.24}$	$48.22_{0.41}$	$62.92_{0.33}$
	Span	$76.73_{0.13}$	$84.96_{0.13}$	$69.54_{0.21}$	$76.11_{0.33}$	$48.59_{0.09}$	$63.15_{0.39}$
	Entity	$76.34_{0.36}$	$84.78_{0.21}$	$69.25_{0.43}$	$75.98_{0.39}$	$48.19_{0.20}$	$62.97_{0.16}$
	Punct.	$\underline{76.85}_{0.09}$	$\underline{85.16}_{0.03}$	$69.41_{0.21}$	$76.14_{0.18}$	$48.58_{0.26}$	$63.14_{0.20}$
	NMG	$\mathbf{76.93}_{\mathbf{0.32}}$	$\mathbf{85.30}_{\mathbf{0.23}}$	$\mathbf{69.98}_{\mathbf{0.37}}$	$\mathbf{76.51}_{\mathbf{0.45}}$	$\mathbf{48.75}_{\mathbf{0.08}}$	$\mathbf{63.55}_{\mathbf{0.14}}$

Table 1: Performance of various masking strategies on the QA tasks. We run the model three times with different random seeds then report the average performances, with standard deviations (subscripts). The numbers in bold fonts denote best scores, and the numbers with underlines denote the second best scores.

Base LM	Model	ChemProt Acc	IMDb Acc
BERT	No PT	$80.40_{0.70}$	$92.28_{0.05}$
	Random	$\underline{81.25}_{0.72}$	$92.45_{0.21}$
	Whole	$80.18_{1.20}$	$\mathbf{92.55}_{\mathbf{0.04}}$
	Span	$78.06_{1.72}$	$92.40_{0.10}$
	Entity	$79.68_{1.32}$	$92.38_{0.13}$
	Punct.	$79.68_{0.30}$	$92.40_{0.18}$
	NMG	$\mathbf{81.66}_{\mathbf{0.37}}$	$\underline{92.53}_{0.03}$

Table 2: Performance of various masking strategies on the text classification tasks. The presentation format is the same as in Table 1.

Base LM: DistilBERT	NewsQA EM	NewsQA F1
NMG	$\mathbf{48.75}_{\mathbf{0.08}}$	$\mathbf{63.55}_{\mathbf{0.14}}$
NMG w/o Self-Play	$48.64_{0.27}$	$63.37_{0.17}$
NMG w/o Continual-Adaptive	$48.52_{0.12}$	$63.06_{0.20}$

Table 3: Ablation Results on the NewsQA dataset.

emrQA. In contrast, such maskings sometimes lead to worse results than random masking on some domains such as in NewsQA. Especially, in the case of DistilBERT, punctuation masking performs better than others. These result suggests that masking complicated words is rather disturbing for the adaptation of small language model.

On the other hand, our NMG learns the optimal masking policy for a given task and domain in an adaptive manner on any language model. In Table 1 and 2, we can see that this adaptive characteristic of our model makes the neural masking results in better or at least comparable performances to the baselines for all tasks. We further analyze the learned masking strategy in Section 5.3.

5.2 Ablation study

Effectiveness of Self-Play We further investigate the effectiveness of self-play by comparing it with the NMG model without self-play, where the model only competes with the random agent. We validate this on the NewsQA dataset. The result in Table 3 shows that the NMG model with self-play obtains better performance than its counterpart without self-play. This result verifies that competing with the opponent neural agent while learning helps the NMG model to learn better policy.

Continual Adaptation We also perform an ablation study of the continual adaptation learning. The result in Table 3 shows that the continual-adaptive masking strategy is significantly effective for the language model adaptation. The result suggests that helpful words for the language model to learn depends on the adaptation degree of it.

5.3 Analysis

Masked Word Statistics To analyze how our model performs, we measure the difference between which kind of word token is masked by both the random and neural policy on the pre-trained checkpoint. For qualitative analysis, we provide examples of masked tokens on the context in Figure 3. As shown in Figure 3, NMG tends to mask highly informative words such as *seminary* or *islam*, which are parts of the answer spans. Furthermore, we analyze the masking behavior of our NMG by performing Part-of-Speech (POS) tagging on the masked words using spaCy[4]. Figure 4 shows

[4]https://spacy.io

Neural (SQuAD) … holy cross (albeit not its official headquarters, which are in rome). its main seminary, moreau seminary, is located on the campus across st. joseph lake from the main building. old college, the oldest building on campus and located near the shore of st. mary lake, houses undergraduate …

Random (SQuAD) … holy cross (albeit not its official headquarters, which are in rome). its main seminary, moreau seminary, is located on the campus across st. joseph lake from the main building. old college, the oldest building on campus and located near the shore of st. mary lake, houses undergraduate …

Question What is the primary seminary of the Congregation of the Holy Cross? **Answer** Moreau Seminary

Neural (NewsQA) … iraqi officials have said the families were frightened by a series of killings and threats by muslim extremists ordering them to convert to islam or face death. fourteen christians have been slain in …

Random (NewsQA) … iraqi officials have said the families were frightened by a series of killings and threats by muslim extremists ordering them to convert to islam or face death. fourteen christians have been slain in …

Question What did extremists order them to do? **Answer** Islam or face death.

Figure 3: Examples of masked tokens using our Neural Mask Generator (Neural) model and random sampling (Random). The tokens colored in red denote masked tokens by the model, and words highlighted with a yellow box is the answer to the given question. For more examples on multiple datasets, **please see Appendix C.**

the six most frequent tags for the words masked out by the random and neural policy. Figure 4 shows that the neural policy masks more words in noun, verb, and proper noun tags than the random policy, suggesting that our NMG model learns that masking such informative words is beneficial to adapt on the NewsQA task with the BERT$_{BASE}$ model as a language model.

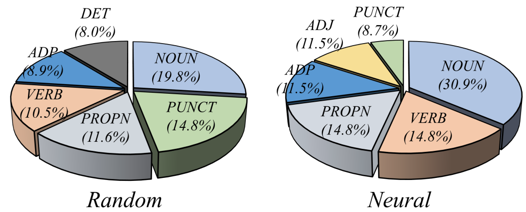

Figure 4: Top-6 POS of Masked Words on NewsQA.

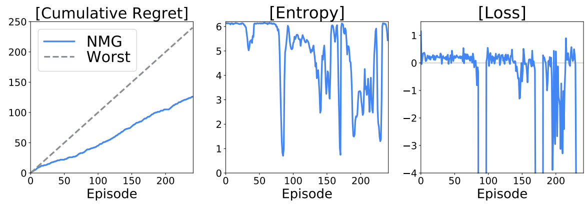

Figure 5: Reinforcement learning curves on NewsQA.

Learning Curves As already known, the RL-based methods often suffer from the instability problem. Therefore, we further analyze the learning curves of the NMG training in this section. In Figure 5, we plot three kinds of learning curves to show the detailed training process. *Cumulative Regret* indicates how many times the neural agent is defeated against the random agent until certain episodes. The grey plot indicates the worst case that the random agent always defeats against the neural agent. *Entropy* indicates the average entropy of policy for states given in a certain episode. Lower entropy means that the policy has a high probability of a few significant actions. *Loss* indicates the RL loss described in the equation 10. We ignore outliers in the loss plot for brevity.

From the entropy and loss plots, we can notice that the policy converges as learning proceeds. However, it seems that such convergence is not continually sustained. From the cumulative regret plot, we can observe that the neural policy still often loses against the random policy, although it is trained for a while. Such instability may come from the difficulty of the exact credit assignment on each action. Otherwise, continuous change of state distribution from the continual adaptive learning may hinder the neural policy's convergence.

Even if the NMG shows the notable results, there is room for improvement on RL in terms of efficiency and stability. We leave it as the future work.

6 Conclusion

We proposed a novel framework which automatically generates an adaptive masking for masked language models based on the given context, for language model adaptation to low-resource domains. To this end, we proposed the *Neural Mask Generator (NMG)*, which is trained with reinforcement learning to mask out words that are helpful for domain adaptation. We performed an empirical study of various rule-based masking strategies on multiple datasets for question answering and text classification tasks, which shows that the optimal masking strategy depends on both the language model and the domain. We then validated NMG against rule-based masking strategies, and the results show that it either outperforms, or obtains comparable performance to the best heuristic. Further qualitative analysis suggests that such good performance comes from its ability to adaptively mask meaningful words for the given task.

Acknowledgments

This work was supported by Samsung Advanced Institute of Technology (SAIT), the Engineering Research Center Program through the National Research Foundation of Korea (NRF) funded by the Korea Government MSIT (NRF2018R1A5A1059921), Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea Government (MSIT) (No.2016-0-00563, Research on Adaptive Machine Learning Technology Development for Intelligent Autonomous Digital Companion, and No.2019-0-00075, Artificial Intelligence Graduate School Program (KAIST)), and a study on the "HPC Support" Project supported by the 'Ministry of Science and ICT' and NIPA.

References

Iz Beltagy, Arman Cohan, and Kyle Lo. 2019. Scibert: Pretrained contextualized embeddings for scientific text. *CoRR*, abs/1903.10676.

Kevin Clark, Minh-Thang Luong, Quoc V. Le, and Christopher D. Manning. 2020. ELECTRA: pre-training text encoders as discriminators rather than generators. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*.

Thomas Degris, Martha White, and Richard S. Sutton. 2012. Linear off-policy actor-critic. In *Proceedings of the 29th International Conference on Machine Learning, ICML 2012, Edinburgh, Scotland, UK, June 26 - July 1, 2012*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. 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, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 4171–4186.

Li Dong, Nan Yang, Wenhui Wang, Furu Wei, Xiaodong Liu, Yu Wang, Jianfeng Gao, Ming Zhou, and Hsiao-Wuen Hon. 2019. Unified language model pre-training for natural language understanding and generation. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, 8-14 December 2019, Vancouver, BC, Canada*, pages 13042–13054.

Chelsea Finn, Pieter Abbeel, and Sergey Levine. 2017. Model-agnostic meta-learning for fast adaptation of deep networks. In *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*, pages 1126–1135.

Luca Franceschi, Paolo Frasconi, Saverio Salzo, Riccardo Grazzi, and Massimiliano Pontil. 2018. Bilevel programming for hyperparameter optimization and meta-learning. In *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, pages 1563–1572.

Michael R. Glass, Alfio Gliozzo, Rishav Chakravarti, Anthony Ferritto, Lin Pan, G. P. Shrivatsa Bhargav, Dinesh Garg, and Avirup Sil. 2019. Span selection pre-training for question answering. *CoRR*, abs/1909.04120.

Jordi Grau-Moya, Felix Leibfried, and Haitham Bou-Ammar. 2018. Balancing two-player stochastic games with soft q-learning. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden*, pages 268–274.

Suchin Gururangan, Ana Marasovic, Swabha Swayamdipta, Kyle Lo, Iz Beltagy, Doug Downey, and Noah A. Smith. 2020. Don't stop pretraining: Adapt language models to domains and tasks. *CoRR*, abs/2004.10964.

Xiaochuang Han and Jacob Eisenstein. 2019. Unsupervised domain adaptation of contextualized embeddings for sequence labeling. 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 2019, Hong Kong, China, November 3-7, 2019*, pages 4237–4247.

Dan Hendrycks and Kevin Gimpel. 2016. Bridging nonlinearities and stochastic regularizers with gaussian error linear units. *CoRR*, abs/1606.08415.

Jeremy Howard and Sebastian Ruder. 2018. Universal language model fine-tuning for text classification. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 1: Long Papers*, pages 328–339.

Lu Jiang, Zhengyuan Zhou, Thomas Leung, Li-Jia Li, and Li Fei-Fei. 2018. Mentornet: Learning data-driven curriculum for very deep neural networks on corrupted labels. In *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, pages 2309–2318.

Mandar Joshi, Danqi Chen, Yinhan Liu, Daniel S. Weld, Luke Zettlemoyer, and Omer Levy. 2019. Spanbert: Improving pre-training by representing and predicting spans. *CoRR*, abs/1907.10529.

Diederik P. Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations,* *ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

Jens Kringelum, Sonny Kim Kjærulff, Søren Brunak, Ole Lund, Tudor I. Oprea, and Olivier Taboureau. 2016. Chemprot-3.0: a global chemical biology diseases mapping. *Database*, 2016.

Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. 2019. ALBERT: A lite BERT for self-supervised learning of language representations. *CoRR*, abs/1909.11942.

Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. 2019. Biobert: a pre-trained biomedical language representation model for biomedical text mining. *CoRR*, abs/1901.08746.

Hanxiao Liu, Karen Simonyan, and Yiming Yang. 2019a. DARTS: differentiable architecture search. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*.

Xiaodong Liu, Pengcheng He, Weizhu Chen, and Jianfeng Gao. 2019b. Multi-task deep neural networks for natural language understanding. In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, pages 4487–4496.

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

Ilya Loshchilov and Frank Hutter. 2019. Decoupled weight decay regularization. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*.

Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng, and Christopher Potts. 2011. Learning word vectors for sentiment analysis. In *The 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference, 19-24 June, 2011, Portland, Oregon, USA*, pages 142–150.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States*, pages 3111–3119.

Volodymyr Mnih, Adrià Puigdomènech Badia, Mehdi Mirza, Alex Graves, Timothy P. Lillicrap, Tim Harley, David Silver, and Koray Kavukcuoglu. 2016. Asynchronous methods for deep reinforcement learning. In *Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016*, pages 1928–1937.

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin A. Riedmiller. 2013. Playing atari with deep reinforcement learning. *CoRR*, abs/1312.5602.

Jiquan Ngiam, Daiyi Peng, Vijay Vasudevan, Simon Kornblith, Quoc V. Le, and Ruoming Pang. 2018. Domain adaptive transfer learning with specialist models. *CoRR*, abs/1811.07056.

Anusri Pampari, Preethi Raghavan, Jennifer J. Liang, and Jian Peng. 2018. emrqa: A large corpus for question answering on electronic medical records. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*, pages 2357–2368.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. 2019. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, 8-14 December 2019, Vancouver, BC, Canada*, pages 8024–8035.

Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2018, New Orleans, Louisiana, USA, June 1-6, 2018, Volume 1 (Long Papers)*, pages 2227–2237.

Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners.

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

Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. 2016. Squad: 100, 000+ questions for machine comprehension of text. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016*, pages 2383–2392.

Sachin Ravi and Hugo Larochelle. 2017. Optimization as a model for few-shot learning. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. Distilbert, a distilled version of BERT: smaller, faster, cheaper and lighter. *CoRR*, abs/1910.01108.

Adam Santoro, Sergey Bartunov, Matthew Botvinick, Daan Wierstra, and Timothy P. Lillicrap. 2016. Meta-learning with memory-augmented neural networks. In *Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016*, pages 1842–1850.

Tom Schaul, John Quan, Ioannis Antonoglou, and David Silver. 2016. Prioritized experience replay. In *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*.

John Schulman, Sergey Levine, Pieter Abbeel, Michael I. Jordan, and Philipp Moritz. 2015. Trust region policy optimization. In *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, pages 1889–1897.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal policy optimization algorithms. *CoRR*, abs/1707.06347.

Chi Sun, Xipeng Qiu, Yige Xu, and Xuanjing Huang. 2019a. How to fine-tune BERT for text classification? In *Chinese Computational Linguistics - 18th China National Conference, CCL 2019, Kunming, China, October 18-20, 2019, Proceedings*, pages 194–206.

Yu Sun, Shuohuan Wang, Yukun Li, Shikun Feng, Xuyi Chen, Han Zhang, Xin Tian, Danxiang Zhu, Hao Tian, and Hua Wu. 2019b. ERNIE: enhanced representation through knowledge integration. *CoRR*, abs/1904.09223.

Yu Sun, Shuohuan Wang, Yukun Li, Shikun Feng, Hao Tian, Hua Wu, and Haifeng Wang. 2019c. ERNIE 2.0: A continual pre-training framework for language understanding. *CoRR*, abs/1907.12412.

Richard S. Sutton and Andrew G. Barto. 2018. *Reinforcement Learning: An Introduction*, second edition. The MIT Press.

Wilson L. Taylor. 1953. "cloze procedure": a new tool for measuring readability. *Journalism Bulletin*, 30(4):415–433.

Sebastian Thrun and Lorien Pratt, editors. 1998. *Learning to Learn*. Kluwer Academic Publishers, Norwell, MA, USA.

Adam Trischler, Tong Wang, Xingdi Yuan, Justin Harris, Alessandro Sordoni, Philip Bachman, and Kaheer Suleman. 2017. Newsqa: A machine comprehension dataset. In *Proceedings of the 2nd Workshop on Representation Learning for NLP, Rep4NLP@ACL 2017, Vancouver, Canada, August 3, 2017*, pages 191–200.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA*, pages 5998–6008.

Oriol Vinyals, Charles Blundell, Tim Lillicrap, Koray Kavukcuoglu, and Daan Wierstra. 2016. Matching networks for one shot learning. In *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, pages 3630–3638.

Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel R. Bowman. 2019. GLUE: A multi-task benchmark and analysis platform for natural language understanding. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*.

Ziyu Wang, Victor Bapst, Nicolas Heess, Volodymyr Mnih, Rémi Munos, Koray Kavukcuoglu, and Nando de Freitas. 2017. Sample efficient actor-critic with experience replay. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*.

Chen-Yu Wei, Yi-Te Hong, and Chi-Jen Lu. 2017. Online reinforcement learning in stochastic games. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA*, pages 4987–4997.

Taesun Whang, Dongyub Lee, Chanhee Lee, Kisu Yang, Dongsuk Oh, and Heuiseok Lim. 2019. Domain adaptive training BERT for response selection. *CoRR*, abs/1908.04812.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, R'emi Louf, Morgan Funtowicz, and Jamie Brew. 2019. Huggingface's transformers: State-of-the-art natural language processing. *ArXiv*.

Zhilin Yang, Zihang Dai, Yiming Yang, Jaime G. Carbonell, Ruslan Salakhutdinov, and Quoc V. Le. 2019. Xlnet: Generalized autoregressive pretraining for language understanding. *CoRR*, abs/1906.08237.

Jinsung Yoon, Sercan Ömer Arik, and Tomas Pfister. 2019. Data valuation using reinforcement learning. *CoRR*, abs/1909.11671.

Linchao Zhu, Sercan Ömer Arik, Yi Yang, and Tomas Pfister. 2019. Learning to transfer learn. *CoRR*, abs/1908.11406.

Barret Zoph and Quoc V. Le. 2017. Neural architecture search with reinforcement learning. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*.

Barret Zoph, Vijay Vasudevan, Jonathon Shlens, and Quoc V. Le. 2018. Learning transferable architectures for scalable image recognition. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*, pages 8697–8710.

A Algorithms

We provide the pseudocode of algorithm for meta-training of the Neural Mask Generator (NMG).

In the case of meta-testing, *InnerLoop* with the full task $\mathcal{T}$, pre-trained language model checkpoint θ, and trained policy π_λ as inputs can be considered as the meta-testing algorithm.

Algorithm 1 NMG Training Algorithm

Initialize random policy $\psi \sim$ *Uniform*
Initialize two neural agents λ, λ^{op} for Self-Play
Initialize replay buffer $\mathcal{D}, \mathcal{D}^{op}$
Arbitrarily split $D_{tr} \to D_{tr}, D_{val}$
Load pre-trained language model θ
while not done **do**
  Randomly Sample $\mathcal{T}'$ from $\mathcal{T}$
  $r_\psi, \mathcal{E}_\psi, \theta(\psi) =$ *InnerLoop*$(\mathcal{T}', \theta, \psi)$
  $r_{op}, \mathcal{E}_{\lambda^{op}}, \theta(\lambda^{op}) =$ *InnerLoop*$(\mathcal{T}', \theta, \pi_{\lambda^{op}})$
  $r, \mathcal{E}_\lambda, \theta(\lambda) =$ *InnerLoop*$(\mathcal{T}', \theta, \pi_\lambda)$
  $\mathcal{A}^{op} \leftarrow \mathcal{E}_\psi, r_\psi, \mathcal{E}, r, \mathcal{E}_{op}, r_{op}$
  $\mathcal{D}^{op}, \lambda^{op} \leftarrow$ *OuterLoop*$(\mathcal{D}^{op}, \mathcal{A}^{op}, \lambda^{op})$
  $\mathcal{A} \leftarrow \mathcal{E}_\psi, r_\psi, \mathcal{E}_{op}, r_{op}, \mathcal{E}, r$
  $\mathcal{D}, \lambda \leftarrow$ *OuterLoop*$(\mathcal{D}, \mathcal{A}, \lambda)$
  $\theta \leftarrow \theta(\lambda)$
end while

Algorithm 2 InnerLoop

Input:
  Task $\mathcal{T}$, LM θ, Policy π
Output:
  Episode buffer $\mathcal{E}$, Accuracy r, LM $\theta(\lambda)$
Initialize episode buffer $\mathcal{E}$
$\hat{\mathbf{S}} = \{\}, \boldsymbol{S}, D_{tr}, D_{val} \leftarrow \mathcal{T}$
for s in $\boldsymbol{S}$ **do**
  Masking s following equation 3, 4
  $\hat{\mathbf{S}} \leftarrow \hat{\mathbf{S}} \cup \hat{s}$, $\mathcal{E} \leftarrow \mathcal{E} \cup (s, a, \pi)$
end for
 # Actually, below two updates are done with mini-batch optimization with multiple steps
$\theta(\lambda) \leftarrow \theta - \nabla_\theta \mathcal{L}_{MLM}(\theta, \lambda; \boldsymbol{S}, \hat{\mathbf{S}})$
$\theta^*(\lambda) \leftarrow \theta(\lambda) - \nabla_{\theta(\lambda)} \mathcal{L}_{train}(\theta(\lambda), \lambda; D_{tr})$
Evaluate θ on D_{val} and acquire r

B Hyperparameters

B.1 Reinforcement Learning (Outer Loop)

We describe detailed hyperparameters in Table 4 for reinforcement learning (RL). We use the prioritized experience replay (Schaul et al., 2016) with

Algorithm 3 OuterLoop

Input:
  $\mathcal{D}$, $\mathcal{E}_\psi$, r_ψ, $\mathcal{E}_{op}$, r_{op}, $\mathcal{E}$, r, Agent λ
Output:
  Replay buffer $\mathcal{D}$, Trained Agent λ
$R \leftarrow sgn(r - r_{op})$
for (s, a) in $\mathcal{E}$ **do**
  if $(s, a) \notin \mathcal{E} \cap \mathcal{E}_{op}$ and $(s, a) \notin \mathcal{E} \cap \mathcal{E}_\psi$ **then**
    $R \leftarrow \min(sgn(r - r_\psi), sgn(r - r_{op}))$
  else if $(s, a) \notin \mathcal{E} \cap \mathcal{E}_{op}$ **then**
    $R \leftarrow sgn(r - r_{op})$
  else if $(s, a) \notin \mathcal{E} \cap \mathcal{E}_\psi$ **then**
    $R \leftarrow sgn(r - r_\psi)$
  else
    continue
  end if
  $\mathcal{D} \leftarrow \mathcal{D} \cup \{(s, a, R, \pi_{old})\}$
end for
Sample a mini-batch $\{(s, a, R, \pi_{old})\}$ from $\mathcal{D}$
$\lambda \leftarrow \lambda + \nabla_\lambda \mathcal{L}$ following equation 8, 9 using sampled replays

exponent value as 1. In addition, we address the concept of susampling (Mikolov et al., 2013) on replay sampling. Specifically, we divide the priority of each replay by the square root of word frequency within the corresponding context. For optimization, we use Adam (Kingma and Ba, 2015) optimizer to train the NMG model and its opponent (Self-Play).

Hyperparameters	Value
Learning Rate	0.0001
Number of Epochs	10
Minibatch Size	64
Replay Buffer Size	50000
Entropy Regularization	0.01
Maximum Episodes	200

Table 4: Hyperparameters for Reinforcement Learning Training (Outer Loop). QA is short for Question Answering and TC is short for Text Classification.

B.2 LM Training in Meta-Train (Inner Loop)

We describe detailed hyperparameters in Table 5 for language model (LM) training in the meta-training. In the case of pre-processing of pre-training dataset, we use the context of triplets in Question Answering and sentences in Text Classification. Especially for emrQA (Pampari et al., 2018), we preprocess it to be same as other QA's formats by removing yes-no and multiple answer type questions. Furthermore, we arbitrarily split a train dataset to a train and a validation set by 8 to 2 since Pampari et al. (2018) do not provide a separate validation set. For optimization, we used AdamW optimizer (Loshchilov and Hutter, 2019), with a linear learning rate scheduler. Each meta-training is done on the two Titan XP or RTX 2080 Ti GPUs and it costs maximum of 2 days in the case of BERT. The batch size is adequately selected according to the size of the data and the model.

B.3 LM Training in Meta-Test

We describe detailed hyperparameters in Table 6 for LM training meta-testing. In the meta-testing, we use same setting described in Section B.2 for pre-processing and optimization. The batch size is also adequately selected according to the size of the data and the model.

Regarding the pre-training epoch and the masking probability p in meta-testing, we use two distinct settings for the baselines and NMG model. For the baselines, we train the LM for 1 epoch with $p = 0.15$ following a conventional setting. However, we observed that the fewer masking with a more pre-training epoch is much more beneficial in the meta-training on the task with long contexts since it makes the NMG evaluate actions more precisely. Therefore, for the NMG model, we train the LM for 3 epochs with $p = 0.05$, following the setting of the meta-training.

B.4 Neural Mask Generator Architecture

The policy network of the NMG model consists of the single self-attention layer and two linear layers. The self-attention layer follows the configuration of the transformer layer of BERT$_{BASE}$. We omit the linear layers of the original transformer (Vaswani et al., 2017) implementation in our architecture. The hidden size of linear layers is 128 and gelu (Hendrycks and Gimpel, 2016) is used as an activation function. The value network also consists of the same linear layers after mean-pooling of word representations. The total number of parameters of the NMG model is approximately 2.5M, which is far smaller than the conventional language model.

B.5 Hyperparameter Searching

For searching proper hyperparameter, we use a manual tuning which tries conventional hyperparameters for the reinforcement learning (RL) and the language model (LM) training. A selection

Table 5: Hyperparameters for LM Training of Meta-Train (Inner Loop).

Hyperparameters	Value
Pre-Training Masking Probability	0.05
Pre-Training Learning Rate	0.00002
Pre-Training Epoch	3
Sampled Pre-training Dataset Size	200
Pre-Training Batch Size	Chosen from {8,16,32}
Maximum sequence length in Pre-Training	512 (QA) or 256 (TC)
Fine-Tuning Learning Rate	0.00003 (QA) or 0.00002 (TC)
Fine-Tuning Epoch	1 (QA) or 5 (TC)
Maximum Training Set Size	1000
Validation set Size	10000
Fine-Tuning Batch Size	Chosen from {8,16,32}

Table 6: Hyperparameters for LM Training of Meta-Test.

Hyperparameters	Value
Pre-Training Masking Probability	Chosen from {0.05, 0.15}
Pre-Training Learning Rate	0.00002
Pre-Training Epoch	Chosen from {1, 3}
Pre-Training Batch Size	Chosen from {12,16,24}
Maximum sequence length in Pre-Training	512 (QA) or 256 (TC)
Fine-Tuning Learning Rate	0.00003 (QA) or 0.00002 (TC)
Fine-Tuning Epoch	2 (QA) or 3 (TC)
Fine-Tuning Batch Size	Chosen from {12,16,32}

criterion depends on the result of the meta-testing. Especially, we set the criterion to F1 score and accuracy for question answering and classification task respectively.

C More Examples

To show the masking strategy from our NMG model, we additionally append additional examples from various datasets used in our experiments.

Context [CLS] architecturally, the school has a catholic character. atop the main building's gold dome is a golden statue of the virgin mary. immediately in front of the main building and facing it, is a copper statue of christ with arms upraised with the legend \"venite ad me omnes\". next to the main building is the basilica of the sacred heart. immediately behind the basilica is the grotto, a marian place of prayer and reflection. It is a replica of the grotto at lourdes, france where the virgin mary reputedly appeared to saint bernadette soubirous in 1858. at the end of the main drive (and in a direct line that connects through 3 statues and the gold dome), is a simple, modern stone statue of mary. [SEP]
Question 1 To whom did the Virgin Mary allegedly appear in 1858 in Lourdes France? **Answer 1** Saint Bernadette Soubirous
Question 2 What sits on top of the Main Building at Notre Dame? **Answer 2** A golden statue of the Virgin Mary
Question 3 What is the Grotto at Notre Dame? **Answer 3** A Marian place of prayer and reflection

Context [CLS] all of notre dame's undergraduate students are a part of one of the five undergraduate colleges at the school or are in the first year of studies program. the first year of studies program was established in 1962 to guide incoming freshmen in their first year at the school before they have declared a major. each student is given an academic advisor from the program who helps them to choose classes that give them exposure to any major in which they are interested. the program also includes a learning resource center which provides time management, collaborative learning, and subject tutoring. this program has been recognized previously, by u.s. news & world report, as outstanding. [SEP]
Question 1 What was created at Notre Dame in 1962 to assist first year students? **Answer 1** The First Year of Studies program
Question 2 Which organization declared the First Year of Studies program at Notre Dame's outstanding? **Answer 2** U.S. News & World Report
Question 3 How many colleges for undergraduates are at Notre Dame? **Answer 3** Five

Context [CLS] the university owns several centers around the world used for international studies and research, conferences abroad, and alumni support. the university has had a presence in london, england, since 1968. since 1998, its london center has been based in the former united university club at 1 suffolk street in trafalgar square. the center enables the colleges of arts & letters, business administration, science, engineering and the law school to develop their own programs in london, as well as hosting conferences and symposia. other global gateways are located in beijing, chicago, dublin, jerusalem and rome. [SEP]
Question 1 In what year did Notre Dame first have a facility in England? **Answer 1** 1968
Question 2 Notre Dame has a center in Beijing, what is it referred to as? **Answer 2** Global Gateways
Question 3 In what year did the Suffolk Street location start to house a Notre Dame facility? **Answer 3** 1998

Figure 6: **SQuAD** Examples of masked tokens using the Neural Mask Generator (NMG). The red mark and yellow box indicates masked tokens by the model and the answer given a question, respectively.

Context [CLS] is awake, alert, in mild to moderate respiratory distress, breathing with a respiratory rate at 34, oximetry noted at 88% on room air, temperature is 102 fahrenheit, pulse is 104, blood pressure is 213/89, and oximetry came up nicely to 93% on 4 liters and 100% on nonrebreather facemask. neck is supple. mucous membranes are dry. lungs reveal slightly decreased breath sounds at the bases; however, there are few crackles noted in the left lower base. heart is tachycardic but regular. the remainder of the exam is unremarkable. laboratory data: white count is noted at 11.6 and hematocrit 28.2. this is close to the patient's baseline. bun and creatinine 69 and 5.8, 5.8 quite elevated for this patient. however, he has been evaluated for recent diagnosis of renal failure by the nephrology service.

(Ellipsis)

he also received nitro paste for preload reduction as well as a foley catheter and lasix, all for his chf. disposition: the patient was admitted to the hospital. ******** not reviewed by attending physician ******** diagnoses: pneumonia, chf, and renal failure. ________________________________ sorensen, saul d: 08/13/79 t: 08/13/79 dictated by: sorensen, saul escription document:4-5119504 bffocus record date: 2080-11-09 manamana impatient rental transplant admit note referring physician: dr. amar jacoby outpatient nephrologist: dr. pao arias reason for admission: living unrelated renal transplant cmv status: donor negative, recipient negative h [SEP]

Question What were the results of the abnormal blood pressure in 2079-08-13?
Answer Blood pressure is 213/89

Context [CLS] and mid thoracic back pain. she has had no motor weakness, abnormal sensation, numbness, tingling or incontinence of bowel or bladder. she denied fevers, chills or sweats. there has been no gi or gu symptoms except for constipation. past medical history: unremarkable. medications: on admission included percocet for pain and multivitamins. allergies: no known drug allergies. family history: the patient 's father died at age 69 with metastatic prostate cancer. the patient 's sister has a history of cervical cancer. social history: the patient is married and lives with her husband. she has no children. she is a nonsmoker. she drinks socially.

(Ellipsis)

there was no scapular tenderness. there was mild lower thoracic tenderness. lungs were clear to auscultation. there were no rales or wheezes. there was no egophony. cardiac exam was tachycardic and regular, normal s1 and s2. There was a ii / vi systolic ejection murmur loudest at the left upper sternal border. there was no heave. abdomen was gravid. bowel sounds were present. the abdomen was tympanitic and nontender . the liver was 11 cm by percussion. there was no liver edge palpable, no spleen tip palpable. rectal revealed normal muscle tone, heme negative stool. extremities revealed 1+ bipedal edema on admission. neurological examination revealed the patient to be alert and oriented times three. cranial nerves ii - xii were intact. motor was 5 [SEP]

Question What lab results does he have that are pertinent to tachycardic diagnosis?
Answer Cardiac exam

Context [CLS] admission date: 2016-03-18 discharge date: 2016-04-01 date of birth: 1929-07-27 sex: m service: surgery allergies: tetracycline attending: samuel a. brown, m.d. chief complaint: transfer from lew is to hallmark health system cmed csru major surgical or invasive procedure: s/p i&d (03-17) s/p debridement (03-18) history of present illness: mr. morton is a 86 yo male transferred from kindred park view specialty hospital of springfield-sample. he developed right foot pain the sunday prior to admission and was seen by his podiatrist, who diagnosed him with gout. he was given colchicine and prednisone. mr. morton then developed more pain and warmth to his right foot later in the week and presented to the youville hospital - woody monica. at this hospital he underwent an i&d (03-17) of a right foot infection and subsequently underwent re-exploration (03-18) for developing necrotizing fascitis. he was transferred to hallmark health system for further care.

(Ellipsis)

muscle compartments, extending from a large area of soft tissue loss seen in the distal lateral foreleg to roughly the mid tibia / fibula, 18 cm distal to the knee joint line. the collection is largest at its most proximal extent, measuring 1.4x0.7cm in the transverse dimension. 2. non-specific myositis involving multiple muscle groups in the foreleg, most severe in the anterior, lateral, and posterior deep compartments. 3. tendinosis of the posterior tibialis and peroneus [SEP]

Question Was developing necrotizing fascitis evaluated before?
Answer Re-exploration

Figure 7: **emrQA** Examples of masked tokens using the Neural Mask Generator (NMG). The red mark and yellow box indicates masked tokens by the model and the answer given a question, respectively.

Context [CLS] new delhi, india (cnn) -- a high court in northern india on friday acquitted a wealthy businessman facing the death sentence for the killing of a teen in a case dubbed "the house of horrors." moninder singh pandher was sentenced to death by a lower court in february. the teen was one of 19 victims -- children and young women -- in one of the most gruesome serial killings in India in recent years. the allahabad high court has acquitted moninder singh pandher, his lawyer sikandar b. kochar told cnn. pandher and his domestic employee surinder koli were sentenced to death in february by a lower court for the rape and murder of the 14-year-old. the high court upheld koli's death sentence, kochar said. the two were arrested two years ago after body parts packed in plastic bags were found near their home in noida, a new delhi suburb. their home was later dubbed a "house of horrors" by the indian media. pandher was not named a main suspect by investigators initially, but was summoned as co-accused during the trial, kochar said. kochar said his client was in australia when the teen was raped and killed. pandher faces trial in the remaining 18 killings and could remain in custody, the attorney said. [SEP]	
Question 1 The court aquitted Moninder Singh Pandher of what crime? **Answer 1** Rape and murder	
Question 2 Who was acquitted? **Answer 2** Moninder Singh Pandher	
Question 3 What was Moninder Singh Pandher acquitted for? **Answer 3** The killing of a teen	

Context [CLS] washington (cnn) -- one of the marines shown in a famous world war ii photograph raising the u.s. flag on iwo jima was posthumously awarded a certificate of u.s. citizenship on tuesday. the marine corps war memorial in virginia depicts strank and five others raising a flag on iwo jima. sgt. michael strank, who was born in czechoslovakia and came to the united states when he was 3, derived u.s. citizenship when his father was naturalized in 1935. however, u.s. citizenship and immigration services recently discovered that strank never was given citizenship papers. at a ceremony tuesday at the marine corps memorial -- which depicts the flag-raising – in arlington, virginia, a certificate of citizenship was presented to strank's younger sister, mary pero. strank and five other men became national icons when an associated press photographer captured the image of them planting an american flag on top of mount suribachi on february 23, 1945. strank was killed in action on the island on march 1, 1945, less than a month before the battle between japanese and u.s. forces there ended. jonathan scharfen, the acting director of cis, presented the citizenship certificate tuesday. he hailed strank as "a true american hero and a wonderful example of the remarkable contribution and sacrifices that immigrants have made to our great republic throughout its history". [SEP]	
Question 1 Where was STrank killed? **Answer 1** On the island	
Question 2 Who was among six who famously raised flag on Iwo Jima? **Answer 2** Sgt. Michael Strank	

Context [CLS] united nations (cnn) -- libyan leader moammar gadhafi on wednesday delivered a lengthy, rambling address in his first appearance before the united nations -- slamming both the u.n. security council and the united states. libyan leader moammar gadhafi addresses the u.n. general assembly on wednesday. he broached conspiracy theories, urged probes into u.s. military activities, and took aim at the structure and the actions of the security council, in a one-hour and 36-minute speech at the u.n. general assembly's annual session. gadhafi called for world unity in confronting various world crises, such as climate change and food shortages, but he aimed his ire at the world body and the united states. dressed in a traditional libyan cap and robe, he elaborated on what he believes is the unfairness of the structure of the u.n. security council, which has five permanent members -- the united states, russia, china, france and britain, each with veto power. in his one hour and 36 minute ramble, gadhafi: [SEP]	
Question 1 What requires unified action? **Answer 1** Confronting various world crises, such as climate change and food shortages	
Question 2 What is unfair? **Answer 2** The structure of the U.N. Security Council	
Question 3 What did Libyan leader tell UN general assembly? **Answer 3** He broached conspiracy theories, urged probes into U.S. military activities, and too k aim at the structure and the actions of the Security Council	

Figure 8: **NewsQA** Examples of masked tokens using the Neural Mask Generator (NMG). The red mark and yellow box indicates masked tokens by the model and the answer given a question, respectively.

Context [CLS] << epidermal growth factor receptor >> inhibitors currently under investigation include the small molecules [[gefitinib]] (iressa, zd1839) and erlotinib (tarceva, osi-774), as well as monoclonal antibodies such as cetuximab (imc-225, erbitux). [SEP]

Ground Truth Inhibitor

Context [CLS] agents that have only begun to undergo clinical evaluation include << ci-1033 >>, an irreversible pan-[[erbb]] tyrosine kinase inhibitor, and pki166 and gw572016, both examples of dual kinase inhibitors (inhibiting epiderma l growth factor receptor and her2). [SEP]

Ground Truth Inhibitor

Context [CLS] << alprenolol >> and bromoacetylalprenololmenthane are competitive slowly reversible antagonists at the [[beta 1-adrenoceptors]] of rat left atria. [SEP]

Ground Truth Antagonist

Context [CLS] discovery and optimization of << anthranilic acid sulfonamides >> as inhibitors of methionine aminopeptidase-2: a structural basis for the reduction of [[albumin]] binding. [SEP]

Ground Truth Downregulartor

Context [CLS] mitiglinide (<< kad-1229 >>), a new anti-diabetic drug, is thought to stimulate [[insulin]] secretion by closing the atp-sensitive j+ (k(atp)) channels in pancreatic beta-cells. [SEP]

Ground Truth Indirect-inhibitor

Context [CLS] when i was a kid of 8, i always watched movies and television that i wasn't supposed to, and this was one of them. It's one of my favorite movies of all time, and it has to be the funniest movie i have ever seen in my life, the acting is excellent, they don't make comedies like this anymore these days (movies that are actually funny and make you laugh without resorting to excrement or some type of vomit-inducing body fluid as in those retarded judd apatow movies starring unfunny non-actors like seth rogen, barf). this movie is a classic with actors who can actually act, and deserve all the accolades. [SEP]

Ground Truth Positive, 1

Context [CLS] steven what have you done you have hit an all new low. it is weird since steven's last film shadow man was directed by the same director who did this trash. shadow man was good this was diabolically bad so bad it wasn't even funny steven is hardly in the movie and feels like he is in a cameo appearance and when he is in the film he is dubbed half the time anyway. as for the action well let's just say the wizard of oz had more action than this trash there is hardly any action in the film and when it does finally arrive it is boring depressing badly shot so called action scenes. seagal hardly kills anyone unlike his over films where he goes one man army ie under siege 1 and 2 and exit wounds. the plot is so confusing with so many plot holes that it doesn't make scenes sometimes. flight of fury better be good what a shame i wasted 5 pounds on this garbage 0 out of ten better luck next time [SEP]

Ground Truth Negative, 0

Context [CLS] one of, if not the worst film to come out of britain in the 80s. this tawdry tale of a middle aged lecher who 'seduces' two teenage scrubbers who babysit for him and his faux-posh wife has nothing to redeem it. in turns gratuitous, puerile, uncouth and unrealistic, this film plumbs the depths as it fails miserably in its attempts to be funny, provocative, intellectual and controversial. perhaps the worst thing about this film is the way the strong cast of george costigan, michelle holmes and siobhan finneran are completely stitched by such a lame script. it's no surprise that this was the late andrea dunbar's only work to make it onto the screen. complete and utter rubbish on every level. [SEP]

Ground Truth Negative, 0

Figure 9: **ChemProt and IMDb** Examples of masked tokens using the Neural Mask Generator (NMG). The red mark indicates masked tokens by the model.