Latent Question Interpretation Through Parameter Adaptation Using Stochastic Neuron

Tetiana Parshakova, Dae-Shik Kim
School of Electrical Engineering, KAIST
ten10@kaist.ac.kr, daeshik@kaist.ac.kr

CEUR-WS.org/Vol-2134/paper07.pdf

Abstract

Many neural network-based question-answering models rely on complex attention mechanisms but they are limited in their ability to capture natural language variability, and to generate diverse and/or reasonable answers. To address this limitation, we propose a module that learns the diversity of the possible interpretations for a given question. In order to identify the possible span of the respective answers, parameters for our question-answering model are adapted using the value of the discrete "interpretation neuron". Additionally, we formulate a semi-supervised variational inference framework and fine-tune the final policy using the rewards from the answer accuracy with the policy gradient optimization. We demonstrate sample answers with induced latent interpretations, suggesting that our model has successfully discovered multiple ways of understanding for a given question. When tested using the Stanford Question Answering Dataset (SQuAD), our model outperformed the current baseline, suggesting the potential validity of the approach described in this work. We open source our implementation in PyTorch[1].

1 Introduction

The task of machine reading comprehension can be defined through paragraph understanding and answering questions that are related to it. It is a crucial task in Natural Language Processing that led to the development of diverse deep learning models. A wide range of these models use the encoder-decoder structure to map sequence (e.g. paragraph and question) to sequence (answer), by encoding the input with a long short-term memory (LSTM) [Hochreiter and Schmidhuber, 1997] into a fixed dimensional vector representation, and then decode the output from that vector with another LSTM [Sutskever *et al.*, 2014]. Variations of this framework have been extensively exploited in the conversation modeling task, where neural networks (NNs) learn the mapping between queries and responses [Vinyals and Le, 2015], as well as machine translation [Bahdanau *et al.*, 2014; Cho *et al.*, 2014], text summarization [Paulus *et al.*, 2017], image captioning [Vinyals *et al.*, 2015b] and more.

SQuAD [Rajpurkar *et al.*, 2016] is a benchmark dataset, that is composed of 100,000+ questions posed by crowd-workers on a set of Wikipedia articles. The answer to each question is a span within a document, and the objective is to predict the starting and ending indices of the answer: a_s, a_e. Hence most models generate two probability distributions over the document in such a way that $P = \{P(a_s), P(a_e|a_s)\}$.

Existing state-of-the-art models attempt to capture the most relevant information for answering the question using complex attention mechanisms. In particular, the key idea lies in multi-layered attention that fuses semantic information from the question into the document. It is achieved by coattention encoders that build richer question-document representation as well as various self-matching structures. These models learn to output distributions over the span indices, and during training get equally penalized for producing answers in distinct positions with the ground truth even if the meaning was similar. Thus, they cannot make basic actions needed to generate natural answers. For example, consider the following triplet (document, question, answer):

> **D**: Newcastle International Airport is located approximately 6 miles (9.7 km) from the city centre on the northern outskirts of the city near Ponteland and is the larger of the two main airports serving the North East. It is connected to the city via the Metro Light Rail system and a journey into Newcastle city centre takes approximately 20 minutes by train.
> **Q**: How far is Newcastle 's airport from the center of town?
> **A**: 6 miles

The span "20 minutes by train" is also a correct answer if the question is interpreted in the perspective of time (which sometimes can be more practical), but since it differs from the ground truth span, the cross entropy loss will discourage this answer. As a consequence, these attention-based discriminative models are limited in their ability to exhibit stochasticity and variability of natural language and to generate diverse yet reasonable answers.

To address this problem, we propose integrating a module that Adapts Parameters through Stochastic Neuron (APSN)

[1] https://github.com/parshakova/apsn

with a basic question-answering model (in our case DrQA, [Chen *et al.*, 2017]) and a training framework for learning a complex distribution of the latent query interpretations during the question answering. The discrete stochastic neuron here represents the interpretation of a question and can be considered as different personas of the answering agent. This stochastic neuron is inferred from the question, and based on its value the central document encoding parameters get adapted to produce an answer for a particular interpretation.

APSN framework employs a discrete latent variable [Mnih and Gregor, 2014], because continuous latent space is harder to interpret and apply for semi-supervised learning environment [Kingma *et al.*, 2014]. The objective is to perform Bayesian inference for the posterior distribution of latent interpretations conditioned on the questions and document sub-spans.

In the framework of variational auto-encoder (VAE), we construct an inference network as the variational approximation of the posterior, and by sampling the interpretation for each question-answer the model is able to learn the interpretation distribution on the SQuAD by optimizing the variational lower bound [Mnih and Gregor, 2014; Miao and Blunsom, 2016; Wen *et al.*, 2017]. To reduce the variance further, we develop the semi-supervised framework by jointly training on the labelled and unlabelled latent interpretations [Kingma *et al.*, 2014].

In order to prevent the mode collapse in selecting only a single interpretation, we introduce a new objective that discourages a cosine similarity and penalizes the feature correlation proximity of original document encoding and document encoding under different latent interpretations. The latter is computed by a mean square error between Gram matrices [Gatys *et al.*, 2015; Gatys *et al.*, 2016]. In addition, after training the model in the semi-supervised variational inference framework, we fine-tune it with a mixed objective that combines traditional cross entropy loss over position of a span with a policy gradient (PG) reinforcement learning [Xiong *et al.*, 2017; Paulus *et al.*, 2017; Li *et al.*, 2016]. In the mixed objective scenario, the latent interpretation is sampled from the prior distribution, and the span distribution is considered to be a policy for PG optimization. We compared the performance of the model with two different scores for obtaining rewards: the F1 score and the exact match (EM).

In summary, our results suggest that the neural variational inference framework is able to detect discrete latent interpretations of a question. Finding various reasonable answers within the same document is important, because it brings a stepping stone towards building a large-scale open-domain QA, where one must first retrieve the few relevant articles and then scan them to identify an answer. By allowing multiple question interpretations, the agent may discover new connections in the knowledge and arrive at more interesting responses. The experimental results also indicate that by introducing a module APSN and training framework to the baseline DrQA, the accuracy of answers on the SQuAD improves. Lastly, the quality of sample answers with induced latent interpretations indicates that the model has successfully discovered multiple ways of understanding the question.

2 Related Work

Among the state-of-the-art end-to-end machine comprehension models on the SQuAD dataset, attention mechanisms play a crucial role.

Bidirectional Attention Flow for Machine Comprehension (BiDAF, [Seo *et al.*, 2016]) was built upon the hierarchical multi-stage architecture. It filters document using the question. Additionally BiDAF symmetrically filters the question using the document, to extract relevant parts of the questions.

The Dynamic Coattention Network (DCN, [Xiong *et al.*, 2016]) uses coattention encoders to fuse the question and paragraph into one representation. It also employs a dynamic decoder that iteratively estimates the start and end indexes using LSTM and a Highway Maxout Network. The extension of DCN, DCN+ [Xiong *et al.*, 2017], introduces the mixed objective of cross entropy loss over span position and self-critical policy learning [Paulus *et al.*, 2017].

The R-Net [Wang *et al.*, 2017] is based on match-LSTMs [Wang and Jiang, 2016] that first incorporate question information into passage representation and then use it for a recurrent self-matching attention. Start and end indices are predicted with the use of pointer networks [Vinyals *et al.*, 2015a].

These models are equipped with a large number of parameters, which is owing to the structure complexity of their attention mechanisms that is loaded with various information pathways and tangled connections between layers. In contrast, DrQA [Chen *et al.*, 2017] is a fairly small and simple model but is powerful enough to achieve a high accuracy on the SQuAD. That is why it was chosen as a baseline model due to being more amenable to fast learning and modifications.

Gradient-based learning has been a key to most neural network based algorithms. The backpropagation [Rumelhart *et al.*, 1986] computes exact gradients when the relationship between the training objective and parameters is continuous and generally smooth. However in many cases it is impossible to apply backpropagation: for example when the model has stochastic neurons, hard non-linearities, discrete sampling operations, or when the objective function is unknown to the agent (like in reinforcement learning). To get a learning signal in such situations one has to construct a gradient estimator.

For models with continuous latent variables the reparametrisation trick is commonly used [Kingma and Welling, 2013] to achieve an unbiased low-variance gradient estimator. While in a discrete latent variable case, advantage actor-critic methods (A2C) give unbiased gradient estimates with reduced variance [Sutton *et al.*, 2000], and a more recent framework RELAX [Grathwohl *et al.*, 2017] that outperforms A2C can be applicable even when no continuous relaxation of discrete random variable is available.

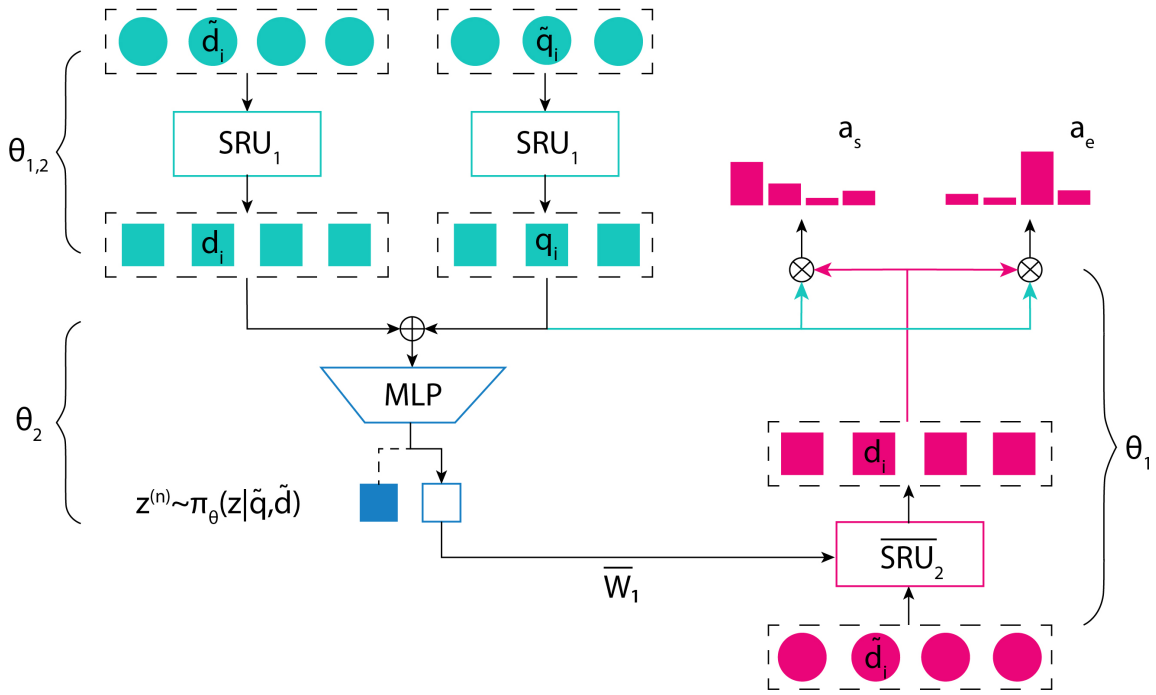

Figure 1: Structure overview of integrated APSN module with DrQA. In this illustration $n_i = 2$ and $z^{(n)} = 1$.

In this work, we investigate the possibilities of modeling the question interpretation distribution during a reading comprehension using the discrete VAE for inference [Mnih and Gregor, 2014], which parametrizes interpretation space through discrete latent variable. This framework is capable of combining different learning paradigms such as semi-supervised learning, reinforcement learning, and sample-based variational inference to bootstrap performance.

3 Baseline and APSN Integration

The APSN module is integrated with the question-answering baseline DrQA. Suppose we are given a document d consisting of m tokens $\{d_1, ..., d_m\}$, a question q consisting of l tokens $\{q_1, ..., q_l\}$, and total n_i of latent question interpretations. We divide model parameters θ into two sets: the policy π parameters θ_2 and all the remaining ones θ_1.

Question Encoding

First, we obtain the question encodings with a multi-layer bidirectional Simple Recurrent Unit (SRU, [Lei and Zhang, 2017]) applied on top of the word embeddings $\widetilde{\mathbf{q}} = \{\widetilde{\mathbf{q}}_1, ..., \widetilde{\mathbf{q}}_l\}$, where $f_{emb}(q_i) = \mathbf{E}(q_i) = \widetilde{\mathbf{q}}_i$:

$$\{\mathbf{q}_1, ..., \mathbf{q}_l\} = \mathrm{SRU}_1\{\widetilde{\mathbf{q}}_1, ..., \widetilde{\mathbf{q}}_l\} \tag{1}$$

Resulting encodings are combined into a question encoding through a parametrized weighted sum $\mathbf{q}^w = \sum_j b_j \mathbf{q}_j$.

Document Encoding

Each token in the document d_i is first preprocessed into a feature vector $\widetilde{\mathbf{d}}_i$ that is comprised of concatenated: word embedding $f_{emb}(d_i) = \mathbf{E}(d_i)$, exact match $f_{em}(d_i) = \mathbb{I}(d_i \in q)$, token features $f_{token}(d_i) = (\mathrm{POS}(d_i), \mathrm{NER}(d_i), \mathrm{TF}(d_i))$, and aligned question embedding $f_{align}(d_i) = \sum_j a_{i,j}\mathbf{E}(q_j)$, as in the original DrQA. To encode the document we apply another recurrent network.

For reference, in the original single-layer SRU linear transformation of the input $\widetilde{\mathbf{d}}$ is performed by grouping matrix multiplication:

$$\mathbf{U}^T = \begin{pmatrix} \mathbf{W} \\ \mathbf{W}_h \\ \mathbf{W}_f \\ \mathbf{W}_r \end{pmatrix} [\widetilde{\mathbf{d}}_1, \widetilde{\mathbf{d}}_2, \cdots, \widetilde{\mathbf{d}}_m] \tag{2}$$

Interpretation Policy

The policy network encodes $\widetilde{\mathbf{q}}$ into $\mathbf{q}^w$ and $\widetilde{\mathbf{d}}$ (word embedding part only) into $\mathbf{d}^w$ with SRU_1. Since empirically we found that it is beneficial to share the question encoding parameters with the prior policy. Then, the latent interpretation z is parametrized by a three layered MLP,

$$\pi_{\theta_2}(z|\widetilde{\mathbf{q}}, \widetilde{\mathbf{d}}) = \sigma(\mathbf{W}_4^T \cdot \mathrm{relu}(\mathbf{W}_3^T \cdot \mathrm{relu}(\mathbf{W}_2^T \cdot \mathrm{relu}(\mathbf{W}_1^T[\mathbf{q}^w \oplus \mathbf{d}^w])))) \tag{3}$$

where σ stands for a softmax, biases are omitted for simplicity, $\mathbf{W}_1, \mathbf{W}_2, \mathbf{W}_3, \mathbf{W}_4$ are trainable parameters, $\oplus$ stands for concatenation. The latent interpretation $z^{(n)} \in \{0, 1, ..., n_i - 1\}$ is sampled from a discrete conditional multinomial distribution $z^{(n)} \sim \pi_{\theta_2}(z|\widetilde{\mathbf{q}}, \widetilde{\mathbf{d}})$.

Parameter Adaptation

In the APSN, the sampled interpretation is used to adapt the central SRU parameters $\mathbb{W}$ from a single layer. For each value of the latent interpretation $z^{(n)} = i$ there is an individual set of weights associated with it, $\overline{\mathbb{W}}_i$. These weights are used to distort the central parameters in order to adapt them for new interpretations:

$$\mathbb{W} = \mathbf{W} \oplus \mathbf{W}_h \oplus \mathbf{W}_f \oplus \mathbf{W}_r \tag{4}$$

$$\overline{\mathbb{W}} = \{\overline{\mathbb{W}}_0, ..., \overline{\mathbb{W}}_{n_i-1}\} \tag{5}$$

In this work, we present multiple methods for combining $\mathbb{W}$ with $\overline{\mathbb{W}}_{z^{(n)}}$ to obtain new parameters $\mathbb{W}^{new}_{z^{(n)}}$, which are being optimized for a particular interpretation:

- Addition: $\mathbb{W}^{new}_{z^{(n)}} = \mathbb{W} + \overline{\mathbb{W}}_{z^{(n)}}$
- Multiplication: $\mathbb{W}^{new}_{z^{(n)}} = \mathbb{W} \odot \sigma(\overline{\mathbb{W}}_{z^{(n)}})$
- Convolutional: $\mathbb{W}^{new}_{z^{(n)}} = \text{CNN}(\mathbb{W}, \overline{\mathbb{W}}_{z^{(n)}})$, which consists of multiple layers of 2D convolutions with a kernel zise of 3×3, ReLU activation between layers and zero padding to keep the original size of the matrix

The above procedure is used to obtain adapted parameters $\mathbb{W}^{new}_{z^{(n)}}$ of a single SRU layer. Similar steps can be followed to find adapted parameters in another layer but with a separate set of perturbation weights. The set of adapted parameters in initial layers of SRU, along with unchanged parameters on the remaining layers, are used in what we call $\overline{\text{SRU}}_2$, to get the encodings of the document information:

$$\{\mathbf{d}_1, ..., \mathbf{d}_m\} = \overline{\text{SRU}}_2\{\widetilde{\mathbf{d}}_1, ..., \widetilde{\mathbf{d}}_m\} \tag{6}$$

Prediction

Similarly to the original model DrQA, we use bilinear term to capture the similarity between $\mathbf{d}_i$ and $\mathbf{q}^w$ and compute the probabilities of each token being start and end of an answer:

$$p_\theta(a_s = i|\widetilde{\mathbf{q}}, \widetilde{\mathbf{d}}) \propto \exp(\mathbf{d}_i \mathbf{W}_s \mathbf{q}^w) \tag{7}$$

$$p_\theta(a_e = i|\widetilde{\mathbf{q}}, \widetilde{\mathbf{d}}) \propto \exp(\mathbf{d}_i \mathbf{W}_e \mathbf{q}^w) \tag{8}$$

4 Training Framework

Inference

To implement sampling from the variational posterior for the given observation, we construct an inference network $q_\phi(z|\widetilde{\mathbf{q}}, a)$ with parameters ϕ as the variational approximation of the posterior distribution $p(z|\widetilde{\mathbf{q}}, a)$ [Mnih and Gregor, 2014; Miao and Blunsom, 2016; Wen *et al.*, 2017]:

$$L = \mathbb{E}_{q_\phi(z|\widetilde{\mathbf{q}},a)}\big[\log p_{\theta_1}(a|z, \widetilde{\mathbf{q}}, \widetilde{\mathbf{d}})\big] - \beta D_{KL}\big[q_\phi(z|\widetilde{\mathbf{q}}, a)||\pi_{\theta_2}(z|\widetilde{\mathbf{q}}, \widetilde{\mathbf{d}})\big] \tag{9}$$

$$\leq \log \sum_z p_{\theta_1}(a|z, \widetilde{\mathbf{q}}, \widetilde{\mathbf{d}})\pi_{\theta_2}(z|\widetilde{\mathbf{q}}, \widetilde{\mathbf{d}}) \tag{10}$$

$$= \log p_\theta(a|\widetilde{\mathbf{q}}, \widetilde{\mathbf{d}}) \tag{11}$$

Note that a coefficient $\beta = 0.1$ scales the learning signal of the KL divergence [Higgins *et al.*, 2016]. Although we are not optimizing the exact variational lower bound, the final goal of learning effective answering model that is based on the question interpretation, is mostly up to the reconstruction error.

The inference network $q_\phi(z|\widetilde{\mathbf{q}}, a)$ is conditioned on the answer embedding, which is a document sub-span $\{\widetilde{\mathbf{d}}_i\}_{i=s}^{e} \subset \{\widetilde{\mathbf{d}}_i\}_{i=1}^{m}$, and the question embeddings $\{\widetilde{\mathbf{q}}_i\}_{i=1}^{l}$, on top of which a recurrent neural network is applied. Then, to obtain a multinomial distribution over the latent interpretations, the concatenation of the resulting hidden units of answer and question encodings is passed through a similar network described in Eq. 3.

During the training we draw N samples $z^{(n)} \sim q_\phi(z|\widetilde{\mathbf{q}}, a)$ independently for computing the gradients. Parameters θ_1 are directly updated by backpropagating the stochastic gradients:

$$\nabla_{\theta_1} L \approx \frac{1}{N} \sum_n \frac{\partial \log p_{\theta_1}(a|z^{(n)}, \widetilde{\mathbf{q}}, \widetilde{\mathbf{d}})}{\partial \theta_1}. \tag{12}$$

Parameters of the prior network θ_2 are trained by mimicking the posterior network:

$$\nabla_{\theta_2} L = \beta \sum_z q_\phi(z|\widetilde{\mathbf{q}}, a) \frac{\partial \log \pi_{\theta_2}(z|\widetilde{\mathbf{q}}, \widetilde{\mathbf{d}})}{\partial \theta_2}. \tag{13}$$

For the parameters ϕ in the posterior network, we firstly define the learning signal as:

$$l(a, z^{(n)}, \widetilde{\mathbf{q}}, \widetilde{\mathbf{d}}) = \log p_{\theta_1}(a|z^{(n)}, \widetilde{\mathbf{q}}, \widetilde{\mathbf{d}}) - \beta\big[\log q_\phi(z^{(n)}|\widetilde{\mathbf{q}}, a) - \log \pi_{\theta_2}(z^{(n)}|\widetilde{\mathbf{q}}, \widetilde{\mathbf{d}})\big]. \tag{14}$$

Then the parameters ϕ are updated by:

$$\nabla_\phi L \approx \frac{1}{N} \sum_n \big[l(a, z^{(n)}, \widetilde{\mathbf{q}}, \widetilde{\mathbf{d}}) - b(\widetilde{\mathbf{q}}, \widetilde{\mathbf{d}})\big] \cdot \frac{\partial \log q_\phi(z^{(n)}|\widetilde{\mathbf{q}}, a)}{\partial \phi}. \tag{15}$$

To reduce the variance in this gradient estimator, which relies on samples from $q_\phi(z|\widetilde{\mathbf{q}}, a)$, we follow the REINFORCE algorithm [Mnih and Gregor, 2014] and introduce a baseline critic network $b(\widetilde{\mathbf{q}}, \widetilde{\mathbf{d}}) = \text{MLP}(\mathbf{q}^w \oplus \mathbf{d}^w)$. During the training, the baseline is updated by minimising the mean square error with the learning signal.

Semi-Supervision

While learning interpretations in a completely unsupervised manner, one major difficulty remains: the high variance of an inference network on the early stages of training. Thus, we adopt a semi-supervised training framework [Kingma *et al.*, 2014]. We used a standard clustering algorithm to generate labels $\hat{z}$ for questions-answer pairs. In this case our training examples are separated into two sets: $(\hat{z}, \widetilde{\mathbf{q}}, \widetilde{\mathbf{d}}, a) \in \mathbb{L}$ and $(\widetilde{\mathbf{q}}, \widetilde{\mathbf{d}}, a) \in \mathbb{U}$, that together produce a joint objective

function:

$$L_{ss} = \alpha\Big[\sum_{(\widetilde{\mathbf{q}},\widetilde{\mathbf{d}},a)\in\mathbb{U}} \mathbb{E}_{q_\phi(z|\widetilde{\mathbf{q}},a)}\big[\log p_{\theta_1}(a|z,\widetilde{\mathbf{q}},\widetilde{\mathbf{d}})\big] - \beta D_{KL}\big[q_\phi(z|\widetilde{\mathbf{q}},a)||\pi_{\theta_2}(z|\widetilde{\mathbf{q}},\widetilde{\mathbf{d}})\big]\Big] + \sum_{(\hat{z},\widetilde{\mathbf{q}},\widetilde{\mathbf{d}},a)\in\mathbb{L}} \log\big[p_{\theta_1}(a|\hat{z},\widetilde{\mathbf{q}},\widetilde{\mathbf{d}})\pi_{\theta_2}(\hat{z}|\widetilde{\mathbf{q}},\widetilde{\mathbf{d}})q_\phi(\hat{z}|\widetilde{\mathbf{q}},a)\big] \quad (16)$$

where α is a balancing parameter between updates from a modified variational bound (Eq. 9) and joint log-likelihood of the fully observed data.

Interpretation Diversity

While training a system in the semi-supervised variational inference framework, the interpretation policy suffers from mode collapse. To prevent that, we maximize a new regularization objective:

$$L_{reg} = \sum_{i=0}^{n_i-1} \Big[-0.1 \cdot \cos(\mathbf{U},\overline{\mathbf{U}}_i) + 0.001 \cdot \mathrm{MSE}(\mathrm{Gram}(\mathbf{U}),\mathrm{Gram}(\overline{\mathbf{U}}_i))\Big] \quad (17)$$

where $\mathbf{U}, \overline{\mathbf{U}}_i$ are linear transformations of the average pooled input to SRU across the time steps (Eq. 2) with and without parameter adaptation respectively, $\cos$ is the cosine similarity, Gram is a Gramian matrix divided by $\mathrm{size}(\mathbf{U})$.

By optimizing this objective, the proximity of document encodings under various interpretations that is obtained by feature correlations (i.e., Gram matrix) and cosine similarity, gets minimized. Gram matrix has remarkable ability of capturing texture information and style [Gatys *et al.*, 2015; Gatys *et al.*, 2016], while cosine similarity is useful for measuring how documents are semantically related. L_{reg} along with the main objective (Eq. 9) aims to find such parameters $\overline{\mathbb{W}}, \mathbb{W}$ that help to make document encodings different in semantics and in style across the latent interpretations, but yet producing the correct answers.

Policy Gradient

After the interpretation policy $\pi_{\theta_2}(z|\widetilde{\mathbf{q}},\widetilde{\mathbf{d}})$ and the answering policy $p_{\theta_1}(a|z,\widetilde{\mathbf{q}},\widetilde{\mathbf{d}})$ are learned, we apply a policy gradient-based reinforcement learning algorithm to fine-tune the parameters θ [Xiong *et al.*, 2017; Paulus *et al.*, 2017; Li *et al.*, 2016]. By sampling $z^{(n)} \sim \pi_{\theta_2}(z|\widetilde{\mathbf{q}},\widetilde{\mathbf{d}})$ and $\hat{a} \sim p_{\theta_1}(a|z^{(n)},\widetilde{\mathbf{q}},\widetilde{\mathbf{d}})$ the system receives a reward $r^{(n)}_{\mathrm{score}}(a,\hat{a})$. The new expected gradient from a mixed objective that includes a cross entropy and a policy gradient is computed as:

$$\nabla_\theta L_{ce+pg} \approx \frac{1}{N}\sum_n \frac{\partial}{\partial\theta}\Big[(1-\gamma)\cdot\log p_\theta(a|z^{(n)},\widetilde{\mathbf{q}},\widetilde{\mathbf{d}}) + \gamma\cdot r^{(n)}_{\mathrm{score}}(a,\hat{a})\log\big[p_\theta(\hat{a}|z^{(n)},\widetilde{\mathbf{q}},\widetilde{\mathbf{d}})\pi_\theta(z^{(n)}|\widetilde{\mathbf{q}},\widetilde{\mathbf{d}})\big]\Big]. \quad (18)$$

We evaluated the performance of the model while different scores were used in computing rewards: F1 and EM (between the ground truth and a predicted span). The final value of $r^{(n)}_{\mathrm{score}}(a,\hat{a})$ was normalized over the batch.

5 Rationale

We will now provide the intuition behind the parameter adaptation and the training framework. Current works in hierarchical reinforcement learning are based on the options framework [Sutton *et al.*, 1999], where a master policy selects among options (sub-policies) to accomplish the final goal. Similarly, our algorithm learns a hierarchical policy, where a master policy $\pi_{\theta_2}(z^{(n)}|\widetilde{\mathbf{q}},\widetilde{\mathbf{d}})$ switches between the interpretation-specific weights $\overline{\mathbb{W}}_{z^{(n)}}$ that fine-tune the shared central weights $\mathbb{W}$ and form a sub-policy (sub-policies correspond to $p_\theta(a|\widetilde{\mathbf{q}},\widetilde{\mathbf{d}})$ with $\mathbb{W}^{new}_{z^{(n)}}$) for a particular interpretation value.

Next, we consider VAE framework. It is used to approximate the posterior distribution over the latent interpretations, so that the system could optimize the variational lower bound of the joint distribution. Hence, by sampling the interpretations for each question and correct document sub-span (answer), the model is able to learn the interpretation distribution on SQuAD. To reduce the variance of an inference network on the early stage of training, we introduce semi-supervised learning signal. While maintaining such framework, the system suffers from mode collapse in the interpretation policy. The mode collapse has been prevented by the use of the interpretation diversity objective. In effect, it led to maximally effective behaviour in the question-answering task.

6 Experiments

Implementation Details

For the word embeddings we use GloVe embeddings pre-trained on the 840B Common Crawl corpus [Pennington *et al.*, 2014]. Each recurrent network is a bidirectional SRU that has 5 layers and the hidden state size 128, as in the baseline DrQA. We apply dropout with $p = 0.8$ to all hidden units of SRU, use mini-batches of size 64. The model is trained by Adamax [Kingma and Ba, 2014] and tuned with early stopping on the validation set. In SQuAD some questions contain several ground truth answers, however during training only a single answer per question was used. We apply the Spacy English language models [Honnibal and Montani, 2017] for tokenization and also generating lemma, part-of-speech, and named entity tags.

The trade-off coefficients α and γ are set to 0.1. The final objective in the semi-supervised variational inference framework is $L_{reg}+L_{ss}$. The parameters from the pre-trained DrQA are used as the initialization for the APSN model. Number of features in the convolutional parameter distortion is set to 64. The baseline critic network is a 3 layered MLP with a hidden size 128. Provided accuracies are obtained on the SQuAD validation set.

To produce self-labelled question clusters for semi-supervised learning of the interpretations, we used Sent2Vec [Pagliardini *et al.*, 2017] to obtain sentence embeddings for

Model	n_i	L_{ce+pg} & r_{EM}		L_{ce+pg} & r_{F1}		$L_{ss} + L_{reg}$	
		F1	EM	F1	EM	F1	EM
APSN_conv3	3	80.60	71.54	80.65	71.51	80.57	71.40
APSN_conv4	3	80.56	71.21	80.64	71.31	80.56	71.32
APSN_conv4	4	80.72	71.58	80.79	71.66	80.75	71.40
APSN_conv5	4	80.45	71.40	80.62	71.43	80.49	71.33
APSN_conv4	5	80.66	71.31	80.59	71.20	80.58	71.23
APSN_conv5	5	**80.98**	**71.48**	**81.11**	**71.80**	**80.91**	71.44
APSN_conv3	8	80.61	71.23	80.67	71.27	80.59	71.38
APSN_conv3	10	80.59	71.69	80.69	71.52	80.67	**71.55**

Table 1: Evaluation on the different number of latent interpretations. In the "APSN_conv5", number 5 depicts the amount of layers in the convolutional architecture for obtaining adapted parameters $\mathbb{W}^{new}$. L_{ce+pg} is the semi-supervised variational inference objective. L_{ce+pg} is a mixed objective: cross entropy and policy gradient, where $r_{\text{F1}}, r_{\text{EM}}$ means that the F1 score and the EM respectively are used in computing rewards.

question-answer pairs, and the KMeans for clustering. The number of labelled interpretations was in range from 30% to 50% across the whole dataset, depending on the value of interpretations n_i.

SQuAD Accuracy

Model	n_i	L_{ss}	
		EM	F1
DrQA	-	70.28	79.50
APSN_add_1	5	70.91	80.32
APSN_mul_1	5	70.87	79.86
APSN_mul_2	10	71.30	80.33
APSN_conv4_1	5	71.29	80.72
APSN_conv5_1	5	**71.88**	**81.09**

Table 2: Evaluation of different architectures for obtaining adapted parameters $\mathbb{W}^{new}$ among modules with additive (add), multiplicative (mul) and convolutional (conv) operations. In the "APSN_conv5_1", the number 1 corresponds to a number of layers in the multi-layered SRU where parameters get adapted; the number 5 depicts an amount of layers in the convolutional architecture.

Empirically obtained evaluation results in Table 2 indicate that convolutional operations for adapting parameters are the most effective with our interpretation policy.

The performance of the model in Table 1 illustrates that the accuracy improves while the number of latent interpretations n_i increases from 3 to 5 and then goes down. Also, it is crucial to find a proper number of layers in convolutional parameter adaptation module individually for each value of n_i. The policy gradient framework consistently improves the accuracy achieved by applying solely semi-supervised variational inference training. The APSN model outperforms the baseline DrQA in all cases.

Interestingly, while the regularization objective is used, the model arrives at its best performance on the SQuAD after being fine-tuned with PG, comparing to the case with a single L_{ss} objective. This is the result of the mode collapse that happens in the latter case, when the central parameters of SRU, $\mathbb{W}$, get adapted only for a single task. In this case $\mathbb{W}$ become insensitive to changes in a latent question interpretation neuron.

Analysis of Samples

The sample answers based on the induced values of latent interpretation are illustrated in Table 3. Among the generated spans, some contain new sequences that do not have a word overlap with the first option of ground truth (that the model was trained with) but yet are the plausible answers (samples #1-3 in Table 3 marked in violet). It was the main goal of the interpretation neuron. Other things to note:

1. While the model was trained only with a single answer per question, it is able to find multiple alternative answers in cases when several different options are included in the gold reference (sample #4-6).
2. We also note that predicted spans of some interpretations are inexplicitly related to the correct answer by the causal relationship (samples #7, #8). In such cases, produced answers contain helpful information about the ground truth even when they are not directly answering the question. It may be a valuable path of future investigations to use such spans as an intermediate step for refining the final answers.
3. A paraphrasing behaviour of a question (sample #9) may be useful in making a question-answering model elicit the best answers [Buck *et al.*, 2017].
4. In 80% of cases, the model finds a span that has an overlap with a true answer but either contains additional words (samples #10-12 answer questions more thoroughly) or is more concise. It can be interpreted as the fact that some people are more talkative while others are laconic.

Table 3: Sample answers from the APSN model with $n_i = 5$ produced by inducing the value of a latent interpretation given the document *D* (here only a part of it is shown) and a question *Q* on SQuAD validation set. In this dataset some questions contain several gold reference answers *A*, however during training only a single answer per question was used. The tuple (1 33.3) represents the value of a latent interpretation 1 and the F1 score 33.3%. In each sample, there are shown two predicted answers, among which the one beside the tuple highlighted in bold was chosen by the policy during testing.

1	D:	A job where there are many workers willing to work a large amount of time (high supply) competing for a job that few require (low demand) will result in a low wage for that job.
	Q:	When there are many workers competing for a few jobs its considered as what?
	A:	['high supply', 'low demand']
	(0	0.0) willing to work a large amount of time
	(1	**33.3)** high supply) competing for a job that few require (low demand
2	D:	ITV Tyne Tees was based at City Road for over 40 years after its launch in January 1959. In 2005 it moved to a new facility on The Watermark business park next to the MetroCentre in Gateshead.
	Q:	Where did ITV Tyne Tees move in 2005?
	A:	['a new facility', 'The Watermark business park']
	(1	**100.0)** The Watermark business park
	(2	0.0) Gateshead
3	D:	It is believed that the civilization was later devastated by the spread of diseases from Europe, such as smallpox.
	Q:	What was believed to be the cause of devastation to the civilization?
	A:	['spread of diseases from Europe']
	(1	**0.0)** smallpox
	(4	100.0) spread of diseases from Europe
4	D:	For Luther, also Christ's life, when understood as an example, is nothing more than an illustration of the Ten Commandments, which a Christian should follow in his or her vocations on a daily basis.
	Q:	What should a Christian follow in his life?
	A:	['Ten Commandments', 'his or her vocations on a daily basis']
	(1	**100.0)** Ten Commandments
	(4	72.7) vocations on a daily basis
5	D:	dynamos in a power house six miles away were repeatedly burned out, due to the powerful high frequency currents set up in them, and which caused heavy sparks to jump through the windings and destroy the insulation
	Q:	What did the sparks do to the insulation?
	A:	['destroy', 'jump through the windings and destroy the insulation']
	(2	100.0) jump through the windings and destroy the insulation
	(3	**100.0)** destroy
6	D:	The situation in New France was further exacerbated by a poor harvest in 1757, a difficult winter, and the allegedly corrupt machinations of François Bigot, the intendant of the territory.
	Q:	What other reason caused poor supply of New France from a difficult winter?
	A:	['poor harvest', 'allegedly corrupt machinations of François Bigot']
	(0	**100.0)** poor harvest
	(1	80.0) the allegedly corrupt machinations of François Bigot, the intendant of the territory
7	D:	As the D-loop moves through the circular DNA, it adopts a theta intermediary form, also known as a Cairns replication intermediate, and completes replication with a rolling circle mechanism.
	Q:	What is a Cairns replication intermediate?
	A:	['a theta intermediary form']
	(0	0.0) a rolling circle mechanism
	(1	**100.0)** a theta intermediary form
8	D:	Research shows that student motivation and attitudes towards school are closely linked to student-teacher relationships. Enthusiastic teachers are particularly good at creating beneficial relations with their students.
	Q:	What type of relationships do enthusiastic teachers cause?
	A:	['beneficial']
	(0	0.0) student-teacher
	(4	**66.7)** beneficial relations
9	D:	Thus, the marginal utility of wealth per person ("the additional dollar") decreases as a person becomes richer.
	Q:	What the marginal utility of wealth per income per person do as that person becomes richer?
	A:	['decreases']
	(0	100) decreases
	(4	**0.0)** the additional dollar

10	D:	Plastoglobuli (...), are spherical bubbles of lipids and proteins about 45–60 nanometers across.	
	Q:	What shape are plastoglobuli?	
	A:	['spherical bubbles', 'spherical']	
	(1	100.0)	spherical
	(2	**36.4)**	spherical bubbles of lipids and proteins about 45–60 nanometers
11	D:	This behaviour started with his learning of the execution of Johann Esch and Heinrich Voes, the first individuals to be martyred by the Roman Catholic Church for Lutheran views	
	Q:	Why were Johann Esch and Heinrich Voes executed by the Catholic Church?	
	A:	['for Lutheran views', 'Lutheran views']	
	(0	100.0)	Lutheran views
	(1	**40.0)**	the first individuals to be martyred by the Roman Catholic Church for Lutheran views
12	D:	the rainforest could be threatened though the 21st century by climate change in addition to deforestation	
	Q:	What are the main threats facing the Amazon rainforest in the current century?	
	A:	['climate change in addition to deforestation']	
	(0	100.0)	climate change in addition to deforestation
	(3	**50.0)**	climate change
13	D:	protesters attempted to enter the test site knowing that they faced arrest (...) they stepped across the "line" and were immediately arrested	
	Q:	What was the result of the disobedience protesting the nuclear site?	
	A:	['arrest', 'were immediately arrested']	
	(1	50.0)	they faced arrest
	(2	**0.0)**	Heistler
14	D:	Oxfam's claims have however been questioned on the basis of the methodology used: by using net wealth (adding up assets and subtracting debts), the Oxfam report, for instance, finds that there are more poor people in the United States and Western Europe than in China (due to a greater tendency to take on debts). Anthony Shorrocks, the lead author of the Credit Suisse report which is one of the sources of Oxfam's data, considers the criticism about debt to be a "silly argument" and "a non-issue . . . a diversion".	
	Q:	Why does Oxfam and Credit Suisse believe their findings are being doubted?	
	A:	['a diversion', 'there are more poor people in the United States and Western Europe than in China']	
	(1	**100.0)**	there are more poor people in the United States and Western Europe than in China
	(2	0.0)	the criticism about debt to be a "silly argument"

Thus, the APSN clearly has multiple modes of understanding the question and, therefore, answering it.

7 Conclusion and Future Works

In this paper we have proposed a training framework and the APSN model for learning question interpretations that help to find various valid answers within the same document. The role of a discrete interpretation neuron is to make the central weights $\mathbb{W}$ more sensitive to a particular interpretation. It allows the model to implement multiple modes of answering, since these weights control document representations that are used to get an answer. An important implication of this study is that when the latent distribution is updated by the rewards from a variational lower bound and then the final policy is fine-tuned by the rewards from the answer accuracy, it provides an effective learning approach for the neural network. The sample answers with induced latent interpretations indicate that the model has successfully discovered multiple ways of understanding the question. Lastly, empirical evaluation results on SQuAD suggest that the integration of the APSN into the baseline DrQA is an effective approach for question answering.

In a fair amount of cases the model produces sub-spans or super-spans, failing to detect multiple question interpretations. Further work needs to be done to establish whether having a single question interpretation is a property of SQuAD, or our language in general.

A single sentence from one language can be mapped to multiple variants in another language, thus another direction worth investigation is to connect APSN with a machine translation model. For that APSN will learn a complex distribution of interpretations in mapping source to target sentences. Then the latent interpretation neuron could be seen as a multiple personas translating a sentence.

The APSN module is integrated with the question-answering model DrQA, however, we believe that other baseline models could bring more insights and better results. It may also be fruitful to apply RELAX framework for computing a low variance gradient estimator for the APSN model instead of semi-supervised variational inference due to its outstanding performance in a game domain. Further research in this area could make multi-interpretation approach a standard component in building the answering system.

Acknowledgments

This work was supported by Brain Korea 21+ Project, BK Electronics and Communications Technology Division, KAIST in 2018. This research was also funded by the Hyundai NGV Company (Project No. G01170378).

References

[Bahdanau *et al.*, 2014] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.

[Buck *et al.*, 2017] Christian Buck, Jannis Bulian, Massimiliano Ciaramita, Andrea Gesmundo, Neil Houlsby, Wojciech Gajewski, and Wei Wang. Ask the right questions: Active question reformulation with reinforcement learning. *arXiv preprint arXiv:1705.07830*, 2017.

[Chen *et al.*, 2017] Danqi Chen, Adam Fisch, Jason Weston, and Antoine Bordes. Reading wikipedia to answer open-domain questions. *arXiv preprint arXiv:1704.00051*, 2017.

[Cho *et al.*, 2014] Kyunghyun Cho, Bart Van Merriënboer, Dzmitry Bahdanau, and Yoshua Bengio. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*, 2014.

[Gatys *et al.*, 2015] Leon Gatys, Alexander S Ecker, and Matthias Bethge. Texture synthesis using convolutional neural networks. In *Advances in Neural Information Processing Systems*, pages 262–270, 2015.

[Gatys *et al.*, 2016] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *Computer Vision and Pattern Recognition (CVPR), 2016 IEEE Conference on*, pages 2414–2423. IEEE, 2016.

[Grathwohl *et al.*, 2017] Will Grathwohl, Dami Choi, Yuhuai Wu, Geoff Roeder, and David Duvenaud. Backpropagation through the void: Optimizing control variates for black-box gradient estimation. *arXiv preprint arXiv:1711.00123*, 2017.

[Higgins *et al.*, 2016] Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. beta-vae: Learning basic visual concepts with a constrained variational framework. 2016.

[Hochreiter and Schmidhuber, 1997] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[Honnibal and Montani, 2017] Matthew Honnibal and Ines Montani. spacy 2: Natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing. *To appear*, 2017.

[Kingma and Ba, 2014] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[Kingma and Welling, 2013] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[Kingma *et al.*, 2014] Diederik P Kingma, Shakir Mohamed, Danilo Jimenez Rezende, and Max Welling. Semi-supervised learning with deep generative models. In *Advances in Neural Information Processing Systems*, pages 3581–3589, 2014.

[Lei and Zhang, 2017] Tao Lei and Yu Zhang. Training rnns as fast as cnns. *arXiv preprint arXiv:1709.02755*, 2017.

[Li *et al.*, 2016] Jiwei Li, Will Monroe, Alan Ritter, Michel Galley, Jianfeng Gao, and Dan Jurafsky. Deep reinforcement learning for dialogue generation. *arXiv preprint arXiv:1606.01541*, 2016.

[Miao and Blunsom, 2016] Yishu Miao and Phil Blunsom. Language as a latent variable: Discrete generative models for sentence compression. *arXiv preprint arXiv:1609.07317*, 2016.

[Mnih and Gregor, 2014] Andriy Mnih and Karol Gregor. Neural variational inference and learning in belief networks. *arXiv preprint arXiv:1402.0030*, 2014.

[Pagliardini *et al.*, 2017] Matteo Pagliardini, Prakhar Gupta, and Martin Jaggi. Unsupervised learning of sentence embeddings using compositional n-gram features. *arXiv preprint arXiv:1703.02507*, 2017.

[Paulus *et al.*, 2017] Romain Paulus, Caiming Xiong, and Richard Socher. A deep reinforced model for abstractive summarization. *arXiv preprint arXiv:1705.04304*, 2017.

[Pennington *et al.*, 2014] Jeffrey Pennington, Richard Socher, and Christopher Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.

[Rajpurkar *et al.*, 2016] Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. Squad: 100,000+ questions for machine comprehension of text. *arXiv preprint arXiv:1606.05250*, 2016.

[Rumelhart *et al.*, 1986] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *nature*, 323(6088):533, 1986.

[Seo *et al.*, 2016] Minjoon Seo, Aniruddha Kembhavi, Ali Farhadi, and Hannaneh Hajishirzi. Bidirectional attention flow for machine comprehension. *arXiv preprint arXiv:1611.01603*, 2016.

[Sutskever *et al.*, 2014] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, pages 3104–3112, 2014.

[Sutton *et al.*, 1999] Richard S Sutton, Doina Precup, and Satinder Singh. Between mdps and semi-mdps: A framework for temporal abstraction in reinforcement learning. *Artificial intelligence*, 112(1-2):181–211, 1999.

[Sutton *et al.*, 2000] Richard S Sutton, David A McAllester, Satinder P Singh, and Yishay Mansour. Policy gradient methods for reinforcement learning with function approximation. In *Advances in neural information processing systems*, pages 1057–1063, 2000.

[Vinyals and Le, 2015] Oriol Vinyals and Quoc Le. A neural conversational model. *arXiv preprint arXiv:1506.05869*, 2015.

[Vinyals *et al.*, 2015a] Oriol Vinyals, Meire Fortunato, and Navdeep Jaitly. Pointer networks. In *Advances in Neural Information Processing Systems*, pages 2692–2700, 2015.

[Vinyals *et al.*, 2015b] Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Erhan. Show and tell: A neural image caption generator. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3156–3164, 2015.

[Wang and Jiang, 2016] Shuohang Wang and Jing Jiang. Machine comprehension using match-lstm and answer pointer. *arXiv preprint arXiv:1608.07905*, 2016.

[Wang *et al.*, 2017] Wenhui Wang, Nan Yang, Furu Wei, Baobao Chang, and Ming Zhou. Gated self-matching networks for reading comprehension and question answering. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 189–198, 2017.

[Wen *et al.*, 2017] Tsung-Hsien Wen, Yishu Miao, Phil Blunsom, and Steve Young. Latent intention dialogue models. *arXiv preprint arXiv:1705.10229*, 2017.

[Xiong *et al.*, 2016] Caiming Xiong, Victor Zhong, and Richard Socher. Dynamic coattention networks for question answering. *arXiv preprint arXiv:1611.01604*, 2016.

[Xiong *et al.*, 2017] Caiming Xiong, Victor Zhong, and Richard Socher. Dcn+: Mixed objective and deep residual coattention for question answering. *arXiv preprint arXiv:1711.00106*, 2017.