Reinforced Multi-task Approach for Multi-hop Question Generation

Deepak Gupta, Hardik Chauhan*, **Akella Ravi Tej***, **Asif Ekbal, and Pushpak Bhattacharyya**
Indian Institute of Technology Patna, India
{deepak.pcs16,asif,pb}@iitp.ac.in
{chauhanhardik23,ravitej.akella}@gmail.com

Abstract

Question generation (QG) attempts to solve the inverse of question answering (QA) problem by generating a natural language question given a *document* and an *answer*. While sequence to sequence neural models surpass rule-based systems for QG, they are limited in their capacity to focus on more than one supporting fact. For QG, we often require multiple supporting facts to generate high-quality questions. Inspired by recent works on multi-hop reasoning in QA, we take up Multi-hop question generation, which aims at generating relevant questions based on supporting facts in the context. We employ multitask learning with the auxiliary task of *answer-aware* supporting fact prediction to guide the question generator. In addition, we also proposed a *question-aware* reward function in a Reinforcement Learning (RL) framework to maximize the utilization of the supporting facts. We demonstrate the effectiveness of our approach through experiments on the multi-hop question answering dataset, HotPotQA. Empirical evaluation shows our model to outperform the single-hop neural question generation models on both automatic evaluation metrics such as BLEU, METEOR, and ROUGE, and human evaluation metrics for quality and coverage of the generated questions.

1 Introduction

In natural language processing (NLP), question generation is considered to be an important yet challenging problem. Given a passage and answer as inputs to the model, the task is to generate a semantically coherent question for the given answer.

In the past, question generation has been tackled using rule-based approaches such as question templates (Lindberg et al., 2013) or utilizing named entity information and predictive argument structures of sentences (Chali and Hasan, 2015). Recently, neural-based approaches have accomplished impressive results (Du et al., 2017; Sun et al., 2018; Kim et al., 2018; Gupta et al., 2019b) for the task of question generation. The availability of large-scale machine reading comprehension datasets such as SQuAD (Rajpurkar et al., 2016), NewsQA (Trischler et al., 2017), MSMARCO (Nguyen et al., 2016) etc. have facilitated research in question answering task. These large-scale datasets have also been used to create the resources (Singh et al., 2019; Gupta et al., 2018a) and evaluate the performance (Asai et al., 2018; Gupta et al., 2018b; Gupta et al., 2019a) of the system in low-resource reading comprehension task. SQuAD dataset itself has been the de facto choice for most of the previous works in question generation. However, 90% of the questions in SQuAD can be answered from a single sentence (Min et al., 2018), hence former QG systems trained on SQuAD are not capable of distilling and utilizing information from multiple sentences. Recently released multi-hop datasets such as `QAngaroo` (Welbl et al., 2018), `ComplexWebQuestions` (Talmor and Berant, 2018) and `HotPotQA` (Yang et al., 2018) are more suitable for building QG systems that required to gather and utilize information across multiple documents as opposed to a single paragraph or sentence.

*Work done during an internship at IIT Patna.

Proceedings of the 28th International Conference on Computational Linguistics, pages 2760–2775
Barcelona, Spain (Online), December 8-13, 2020

In multi-hop question answering, one has to reason over multiple relevant sentences from differ-

Document: A few sects, such as the Bishnoi, lay special emphasis on the conservation of particular species, such as the antelope. ***(ii)*** ...
Question$_{SHQ}$**:** *Who lay special emphasis on conservation of particular species ?*
Document (1): ***(i)*** Stig Lennart Blomqvist (born 29 July 1946) is a Swedish rally driver. ***(ii)*** ... ***(iii)*** Driving an Audi Quattro for the Audi factory team, Blomqvist won the World Rally Championship drivers' title in 1984 and finished runner-up in 1985. **Document (2):** ***(i)*** The Audi Quattro is a road and rally car, produced by the German automobile manufacturer Audi, part of the Volkswagen Group. ***(ii)*** ... **Question**$_{MHQ}$**:** *Which car produced by German automobile manufacturer, was driven by Stig Lennart Blomqvist?*

Table 1: An example of Single-hop question (SHQ) from the SQuAD dataset and a Multi-hop Question (MHQ) from the HotPotQA dataset. The relevant sentences and answer required to form the question are highlighted in blue and red respectively.

ent paragraphs to answer a given question. We refer to these relevant sentences as supporting facts in the context. Hence, we frame *Multi-hop question generation* as the task of generating the question conditioned on the information gathered from reasoning over all the supporting facts across multiple paragraphs/documents. Since this task requires assembling and summarizing information from multiple relevant documents in contrast to a single sentence/paragraph, therefore, it is more challenging than the existing single-hop QG task. Further, the presence of irrelevant information makes it difficult to capture the supporting facts required for question generation. The explicit information about the supporting facts in the document is not often readily available, which makes the task more complex. In this work, we provide an alternative to get the supporting facts information from the document with the help of multi-task learning. Table 1 gives sample examples from SQuAD and HotPotQA dataset. It is cleared from the example that the single-hop question is formed by focusing on a single sentence/document and answer, while in multi-hop question, multiple supporting facts from different documents and answer are accumulated to form the question.

Multi-hop QG has real-world applications in several domains, such as education, chatbots, etc. The questions generated from the multi-hop approach will inspire critical thinking in students by encouraging them to reason over the relationship between multiple sentences to answer correctly. Specifically, solving these questions requires higher-order cognitive-skills (e.g., applying, analyzing). Therefore, forming challenging questions is crucial for evaluating a student's knowledge and stimulating self-learning. Similarly, in goal-oriented chatbots, multi-hop QG is an important skill for chatbots, e.g., in initiating conversations, asking and providing detailed information to the user by considering multiple sources of information. In contrast, in a single-hop QG, only single source of information is considered while generation.

In this paper, we propose to tackle Multi-hop QG problem in two stages. In the first stage, we learn supporting facts aware encoder representation to predict the supporting facts from the documents by jointly training with question generation and subsequently enforcing the utilization of these supporting facts. The former is achieved by sharing the encoder weights with an *answer-aware* supporting facts prediction network, trained jointly in a multi-task learning framework. The latter objective is formulated as a *question-aware* supporting facts prediction reward, which is optimized alongside supervised sequence loss. Additionally, we observe that multi-task framework offers substantial improvements in the performance of question generation and also avoid the inclusion of noisy sentences information in generated question, and reinforcement learning (RL) brings the complete and complex question to otherwise maximum likelihood estimation (MLE) optimized QG model.

Our main contributions in this work are: **(i).** We introduce the problem of multi-hop question generation and propose a multi-task training framework to condition the shared encoder with supporting facts information. **(ii).** We formulate a novel reward function, multihop-enhanced reward via *question-aware* supporting fact predictions to enforce the maximum utilization of supporting facts to generate a question; **(iii).** We introduce an automatic evaluation metric to measure the coverage of supporting facts in the generated question. **(iv).** Empirical results show that our proposed method outperforms the current state-of-the-art single-hop QG models over several automatic and human evaluation metrics on the HotPotQA dataset.

2 Related Work

Question generation literature can be broadly divided into two classes based on the features used for generating questions. The former regime consists of rule-based approaches (Heilman and Smith, 2010; Chali and Hasan, 2015) that rely on human-designed features such as named-entity information, etc. to leverage the semantic information from a context for question generation. In the second category, question generation problem is treated as a sequence-to-sequence (Sutskever et al., 2014) learning problem, which involves automatic learning of useful features from the context by leveraging the sheer volume of training data. The first neural encoder-decoder model for question generation was proposed in Du et al. (2017). However, this work does not take the answer information into consideration while generating the question. Thereafter, several neural-based QG approaches (Sun et al., 2018; Zhao et al., 2018; Chen et al., 2018) have been proposed that utilize the answer position information and copy mechanism. Wang et al. (2017a) and Guo et al. (2018) demonstrated an appreciable improvement in the performance of the QG task when trained in a multi-task learning framework.

The model proposed by Seo et al. (2017b) and Weissenborn et al. (2017) for single-document QA experience a significant drop in accuracy when applied in multiple documents settings. This shortcoming of single-document QA datasets is addressed by newly released multi-hop datasets (Welbl et al., 2018; Talmor and Berant, 2018; Yang et al., 2018) that promote multi-step inference across several documents. So far, multi-hop datasets have been predominantly used for answer generation tasks (Seo et al., 2017a; Tay et al., 2018; Zhang et al., 2018). Our work can be seen as an extension to single hop question generation where a non-trivial number of supporting facts are spread across multiple documents.

3 Proposed Approach

Problem Statement: In multi-hop question generation, we consider a document list L with n_L documents, and an m-word answer A. Let the total number of words in all the documents $D_i \in L$ combined be N. Let a document list L contains a total of K candidate sentences $CS = \{S_1, S_2, \ldots, S_K\}$ and a set of supporting facts[1] SF such that $SF \in CS$. The answer $A = \{w_{D_k^{a_1}}, w_{D_k^{a_2}}, \ldots, w_{D_k^{a_m}}\}$ is an m-length text span in one of the documents $D_k \in L$. Our task is to generate an n_Q-word question sequence $\hat{Q} = \{y_1, y_2, \ldots, y_{n_Q}\}$ whose answer is based on the supporting facts SF in document list L. Our proposed model for multi-hop question generation is depicted in Figure 1.

3.1 Multi-Hop Question Generation Model

In this section, we discuss the various components of our proposed Multi-Hop QG model. Our proposed model has four components *(i). Document and Answer Encoder* which encodes the list of documents and answer to further generate the question, *(ii). Multi-task Learning* to facilitate the QG model to automatically select the supporting facts to generate the question, *(iii). Question Decoder*, which generates questions using the pointer-generator mechanism and *(iv). MultiHop-Enhanced QG* component which forces the model to generate those questions which can maximize the supporting facts prediction based reward.

3.1.1 Document and Answer Encoder

The encoder of the Multi-Hop QG model encodes the answer and documents using the layered Bi-LSTM network.

[1]It is only used to train the network. At the time of testing, network predict the supporting facts to be used for question generation.

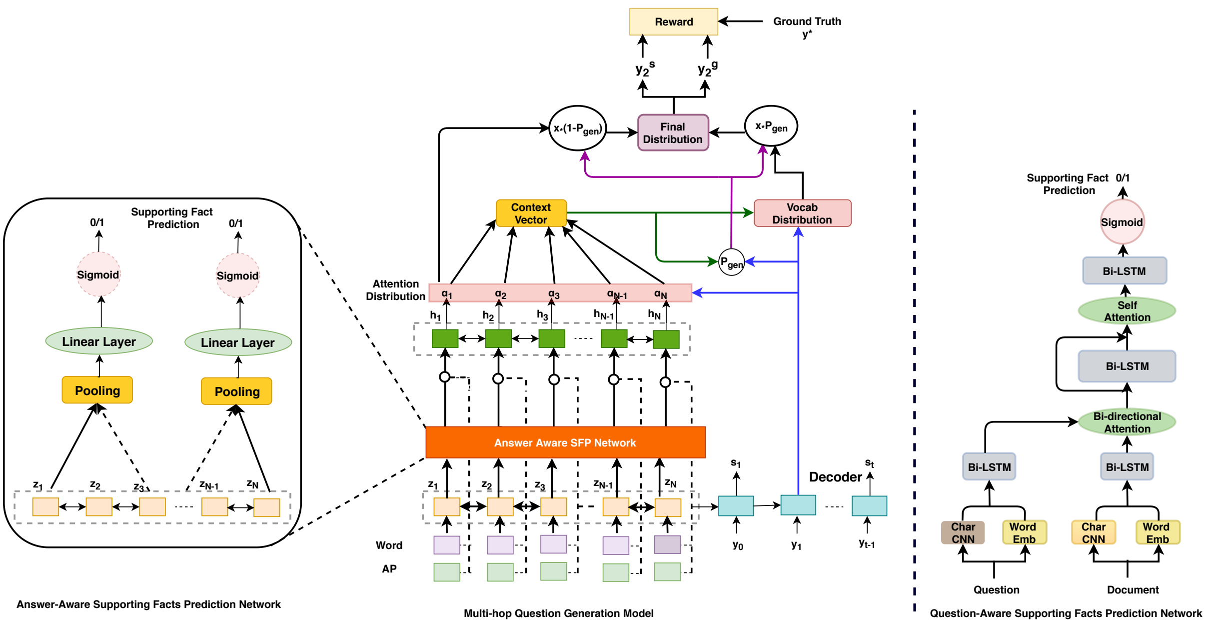

Figure 1: Architecture of our proposed Multi-hop QG network. The inputs to the model are the document word embeddings and answer position (**AP**) features. Question generation and answer-aware supporting facts prediction model (**left**) jointly train the shared document encoder (Bi-LSTM) layer. The image on the **right** depicts our question-aware supporting facts prediction network, which is our MultiHop-Enhanced Reward function. The inputs to this model are the generated question (output of multi-hop QG network) and a list of documents.

Answer Encoding: We introduce an answer tagging feature that encodes the relative position information of the answer in a list of documents. The answer tagging feature is an N length list of vector of dimension d_1, where each element has either a tag value of 0 or 1. Elements that correspond to the words in the answer text span have a tag value of 1, else the tag value is 0. We map these tags to the embedding of dimension d_1. We represent the answer encoding features using $\{a_1, \ldots, a_N\}$.

Hierarchical Document Encoding: To encode the document list L, we first concatenate all the documents $D_k \in L$, resulting in a list of N words. Each word in this list is then mapped to a d_2 dimensional word embedding $u \in \mathbb{R}^{d_2}$. We then concatenate the document word embeddings with answer encoding features and feed it to a bi-directional LSTM encoder $\{LSTM^{fwd}, LSTM^{bwd}\}$.

$$z_t = LSTM(z_{t-1}, [u_t, a_t]) \tag{1}$$

We compute the forward hidden states $\vec{z}_t$ and the backward hidden states $\overleftarrow{z}_t$ and concatenate them to get the final hidden state $z_t = [\vec{z}_t \oplus \overleftarrow{z}_t]$. The answer-aware supporting facts predictions network (will be introduced shortly) takes the encoded representation as input and predicts whether the candidate sentence is a supporting fact or not. We represent the predictions with $p_1, p_2, \ldots, p_K$. Similar to answer encoding, we map each prediction p_i with a vector v_i of dimension d_3.

A candidate sentence S_i contains the n_i number of words. In a given document list L, we have K candidate sentences such that $\sum_{i=1}^{i=K} n_i = N$. We generate the supporting fact encoding $sf_i \in \mathbb{R}^{n_i \times d_3}$ for the candidate sentence S_i as follows:

$$sf^i = e_{n_i} v_i^{\mathrm{T}} \tag{2}$$

where $e_{n_i} \in \mathbb{R}^{n_i}$ is a vector of 1s. The rows of sf_i denote the supporting fact encoding of the word present in the candidate sentence S_i. We denote the supporting facts encoding of a word w_t in the document list L with $s_t \in \mathbb{R}^{d_3}$. Since, we also deal with the answer-aware supporting facts predictions in a multi-task setting, therefore, to obtain a supporting facts induced encoder representation, we introduce another Bi-LSTM layer.

$$h_t = LSTM(h_{t-1}, [z_t, u_t, a_t, s_t]) \tag{3}$$

Similar to the first encoding layer, we concatenate the forward and backward hidden states to obtain the final hidden state representation.

3.1.2 Multi-task Learning for Supporting Facts Prediction

We introduce the task of *answer-aware supporting facts prediction* to condition the QG model's encoder with the supporting facts information. Multi-task learning facilitates the QG model to automatically select the supporting facts conditioned on the given answer. This is achieved by using a multi-task learning framework where the *answer-aware* supporting facts prediction network and Multi-hop QG share a common document encoder (Section 3.1.1). The network takes the encoded representation of each candidate sentence $S_i \in CS$ as input and sentence-wise predictions for the supporting facts. More specifically, we concatenate the first and last hidden state representation of each candidate sentence from the encoder outputs and pass it through a fully-connected layer that outputs a Sigmoid probability for the sentence to be a supporting fact. The architecture of this network is illustrated in Figure 1 (**left**). This network is then trained with a binary cross entropy loss and the ground-truth supporting facts labels:

$$\mathcal{L}_{sp} = -\frac{1}{N}\sum_{j=1}^{N}\sum_{i=1}^{n_j} \delta_{y_i^j=1}\log(p_i^j) + (1-\delta_{y_i^j\neq 1})\log(1-p_i^j) \tag{4}$$

where N is the number of document list, S the number of candidate sentences in a particular training example, δ_i^j and p_i^j represent the ground truth supporting facts label and the output Sigmoid probability, respectively.

3.1.3 Question Decoder

We use a LSTM network with global attention mechanism (Luong et al., 2015) to generate the question $\hat{Q} = \{y_1, y_2, \ldots, y_m\}$ one word at a time. We use copy mechanism (See et al., 2017; Gulcehre et al., 2016) to deal with rare or unknown words. At each timestep t,

$$s_t = LSTM(s_{t-1}, y_{t-1}) \tag{5}$$

The attention distribution α_t and context vector c_t are obtained using the following equations:

$$\begin{aligned} e_{t,i} &= s_t^T * h_i \\ \alpha_{t,i} &= \frac{\exp(e_{t,i})}{\sum_{j=1}^{N}\exp(e_{t,j})} \\ c_t &= \sum_{i=1}^{N}\alpha_{t,i}h_i \end{aligned} \tag{6}$$

The probability distribution over the question vocabulary is then computed as,

$$P_{vocab} = \text{softmax}(\tanh(\mathbf{W_q} * [c_t \oplus s_t])) \tag{7}$$

where $\mathbf{W_q}$ is a weight matrix. The probability of picking a word (generating) from the fixed vocabulary words, or the probability of not copying a word from the document list L at a given timestep t is computed by the following equation:

$$P_{gen} = 1 - \sigma(\mathbf{W_a}c_t + \mathbf{W_b}s_t) \tag{8}$$

where, $\mathbf{W_a}$ and $\mathbf{W_b}$ are the weight matrices and σ represents the Sigmoid function. The probability distribution over the words in the document is computed by summing over all the attention scores of the corresponding words:

$$P_{copy}(w) = \sum_{i=1}^{N}\alpha_{t,i} * \mathbf{1}\{w == w_i\} \tag{9}$$

where $\mathbf{1}\{w == w_i\}$ denotes the vector of length N having the value 1 where $w == w_i$, otherwise 0. The final probability distribution over the dynamic vocabulary (document and question vocabulary) is calculated by the following:

$$P(w) = P_{gen} * P_{vocab}(w) + (1 - P_{gen}) * P_{copy}(w) \tag{10}$$

3.1.4 MultiHop-Enhanced QG

We introduce a reinforcement learning based reward function and sequence training algorithm to train the RL network. The proposed reward function forces the model to generate those questions which can maximize the reward.

MultiHop-Enhanced Reward (MER): Our reward function is a neural network, we call it *Question-Aware Supporting Fact Prediction* network. We train our neural network based reward function for the supporting fact prediction task on `HotPotQA` dataset. This network takes as inputs the list of documents L and the generated question $\hat{Q}$, and predicts the supporting fact probability for each candidate sentence. This model subsumes the latest technical advances of question answering, including character-level models, self-attention (Wang et al., 2017b), and bi-attention (Seo et al., 2017b). The network architecture of the supporting facts prediction model is similar to Yang et al. (2018), as shown in Figure 1 (**right**). For each candidate sentence in the document list, we concatenate the output of the self-attention layer at the first and last positions, and use a binary linear classifier to predict the probability that the current sentence is a supporting fact. This network is pre-trained on HotPotQA dataset using binary cross-entropy loss.

For each generated question, we compute the F1 score (as a reward) between the ground truth supporting facts and the predicted supporting facts. This reward is supposed to be carefully used because the QG model can cheat by greedily copying words from the supporting facts to the generated question. In this case, even though high MER is achieved, the model loses the question generation ability. To handle this situation, we regularize this reward function with additional Rouge-L reward, which avoids the process of greedily copying words from the supporting facts by ensuring the content matching between the ground truth and generated question. We also experiment with BLEU as an additional reward, but Rouge-L as a reward has shown to outperform the BLEU reward function (Please see the **Appendix** for the results).

Adaptive Self-critical Sequence Training: We use the REINFORCE (Williams, 1992) algorithm to learn the policy defined by question generation model parameters, which can maximize our expected rewards. To avoid the high variance problem in the REINFORCE estimator, self-critical sequence training (SCST) (Rennie et al., 2017) framework is used for sequence training that uses greedy decoding score as a baseline. In SCST, during training, two output sequences are produced: y^s, obtained by sampling from the probability distribution $P(y^s_t|y^s_1, \ldots, y^s_{t-1}, \mathcal{D})$, and y^g, the greedy-decoding output sequence. We define $r(y, y^*)$ as the reward obtained for an output sequence y, when the ground truth sequence is y^*. The SCST loss can be written as,

$$\mathcal{L}^{scst}_{rl} = -(r(y^s, y^*) - r(y^g, y^*)) * R \tag{11}$$

where, $R = \sum_{t=1}^{n'} \log P(y^s_t|y^s_1, \ldots, y^s_{t-1}, \mathcal{D})$. However, the greedy decoding method only considers the single-word probability, while the sampling considers the probabilities of all words in the vocabulary. Because of this the greedy reward $r(y^g, y^*)$ has higher variance than the Monte-Carlo sampling reward $r(y^s, y^*)$, and their gap is also very unstable. We experiment with the SCST loss and observe that greedy strategy causes SCST to be unstable in the training progress. Towards this, we introduce a weight history factor similar to (Zhu et al., 2018). The history factor is the ratio of the mean sampling reward and mean greedy strategy reward in previous k iterations. We update the SCST loss function in the following way:

$$\mathcal{L}_{rl} = -\big(r(y^s, y^*) - \alpha \frac{\sum_{i=t-h+1}^{i=t} r_i(y^s, y^*)}{\sum_{i=t-h+1}^{i=t} r_i(y^g, y^*)} r(y^g, y^*)\big) \times R \tag{12}$$

where α is a hyper-parameter, t is the current iteration, h is the history determines, the number of previous rewards are used to estimate. The denominator of the history factor is used to normalize the current greedy reward $r(y^g, y^*)$ with the mean greedy reward of previous h iterations. The numerator of the history factor ensures the greedy reward has a similar magnitude with the mean sample reward of previous h iterations.

Model	BLEU-1	BLEU-2	BLEU-3	BLEU-4	METEOR	ROUGE-L	SF Coverage
s2s (Du et al., 2017)	34.98	22.55	16.79	13.25	17.58	33.75	59.61
s2s+copy	31.86	22.47	17.70	14.63	19.47	30.45	60.48
s2s+answer	39.63	27.35	21.00	16.83	17.58	33.75	61.26
NQG (Zhou et al., 2017)	39.82	29.24	23.45	19.55	21.39	36.63	61.55
ASs2s (Kim et al., 2018)	39.08	29.06	23.45	19.66	22.84	36.98	64.22
Max-out Pointer (Zhao et al., 2018)	42.58	30.91	24.61	20.39	20.36	35.31	63.93
Semantic-Reinforced (Zhang and Bansal, 2019)	44.07	32.72	26.18	21.69	**23.61**	39.40	68.74
SharedEncoder-QG (Ours)	41.72	30.75	24.72	20.64	22.01	37.18	65.46
MTL-QG (Ours)	44.17	32.34	25.74	21.28	21.21	37.55	70.11
Proposed Model	**46.80**	**34.94**	**28.21**	**23.57**	22.88	**39.68**	**74.37**

Table 2: Performance comparison between proposed approach and state-of-the-art QG models on the test set of `HotPotQA`. Here ***s2s***: sequence-to-sequence, ***s2s+copy***: s2s with copy mechanism (See et al., 2017), ***s2s+answer***: s2s with answer encoding.

4 Experimental Setup

4.1 Network Training

We train our multi-task learning based question generator model (*MTL-QG*) using "teacher forcing" algorithm (Williams and Zipser, 1989) to minimize the negative log-likelihood of the model on the training data. With $y^* = \{y_1^*, y_2^*, \dots, y_m^*\}$ as the ground-truth output sequence for a given input sequence D, the maximum-likelihood training objective can be written as,

$$\mathcal{L}_{ml} = -\sum_{t=1}^{m} \log P(y_t^* | y_1^*, \dots, y_{t-1}^*, \mathcal{D}) \tag{13}$$

Since, we are jointly training the supporting-fact prediction network with question generation network then the total loss for the *MTL-QG* network is $\mathcal{L}_{ml} + \beta\mathcal{L}_{sp}$. We use a mixed-objective learning function (Wu et al., 2016; Paulus et al., 2018) to train the final network:

$$\mathcal{L}_{mixed} = \gamma_1\mathcal{L}_{rl} + \gamma_2\mathcal{L}_{ml} + \gamma_3\mathcal{L}_{sp}, \tag{14}$$

where γ_1, γ_2, and γ_3 correspond to the weights of $\mathcal{L}_{rl}$, $\mathcal{L}_{ml}$, and $\mathcal{L}_{sp}$, respectively. We initially train the QG model with multi-task learning (*MTL-QG*) setup. We choose the best *MTL-QG* model which achieves the highest BLEU score on the development dataset. The multihop-enhanced reward based QG model is trained by initializing the best *MTL-QG* model parameters.

4.2 Dataset

We use the `HotPotQA` (Yang et al., 2018) dataset to evaluate our methods. This dataset consists of over 113k Wikipedia-based question-answer pairs, with each question requiring multi-step reasoning across multiple supporting documents to infer the answer. While there exists other multi-hop datasets (Welbl et al., 2018; Talmor and Berant, 2018), only `HotPotQA` dataset provides the sentence-level labels to locate the supporting facts in the list of documents. The ground-truth information of the supporting facts facilitates stronger supervision for tracing the multi-step reasoning chains across the documents used to link the question with the answer. Our approach utilizes the ground-truth supporting facts information (only at the time of training) to form a better input representation and reinforcing desired behavior through multi-task learning and adaptive self-critical RL framework, respectively.

Each question in `HotPotQA` is associated with 10 documents and the span information of the answer and supporting facts in these documents. However, only two of these documents effectively contain all the supporting facts and the ground-truth answer. The average number of supporting facts associated

Document (1): (a) after bedřich smetana, he was the second czech composer to achieve worldwide recognition	
Document (2): (a) concert at the end of summer (czech : koncert na konci léta) is 1980 czechoslovak historical film	
Target Answer: bedřich smetana	
Reference: *which czech composer achieved worldwide recognition before the subject of " concert at the end of summer " ?* **with only Rouge-L reward:** who was the composer of the composer of concert at the end of summer ?	
with Rouge-L and MER: what was the **second czech composer** to achieve worldwide recognition for the composer of the concert at the end of summer ?	
Document (1): (a) seedley railway station is a disused station located in the seedley area of pendleton , salford , on the liverpool and manchester railway	
Document (2): (a) pendleton is an inner city area of salford in greater manchester , england	
Target Answer: england	
Reference: *seedley railway station is a disused station located in the seedley area of pendleton , is an inner city area of salford in greater manchester , in which country ?*	
with only Rouge-L reward: : seedley railway station is located in a city area of salford in what country ?	
with Rouge-L and MER: seedley railway station is a **disused station located** in a city area of salford in greater manchester , in which country ?	

Table 3: Sample questions, where our proposed reward MER based model generating better questions than only Rouge-L reward. The additional included information in the generated questions are shown in Green

with a question is 2.38 (these are all present in the 2, out of 10 documents that contain answer and all the supporting facts). The average length of a question and a document in HotPotQA are 21.82 and 198.3, respectively. In the pre-processing stage, we remove all the documents that do not contain an answer or supporting facts. Further, we remove all the comparison based question-answer pairs that have a '*yes*' or '*no*' answer. We then combine the training set $(90, 564)$ and development set $(7, 405)$ and randomly split the resulting data, with 80% for training, 10% for development, 10% for testing. We call this split as training, development and test dataset of HotPotQA. We ensure that the proportion of difficulty level (easy, medium, hard) of questions was nearly uniform in the train, dev, and test sets.

4.3 Hyperparameters

In our experiments, we use the same vocabulary for both the encoder and decoder. Our vocabulary consists of the top 50,000 frequent words from the training data. We use the development dataset for hyper-parameter tuning. Pre-trained GloVe embeddings (Pennington et al., 2014) of dimension 300 are used in the document encoding step. The hidden dimension of all the LSTM cells is set to 512. Answer tagging features and supporting facts position features are embedded to 3-dimensional vectors. The dropout (Srivastava et al., 2014) probability p is set to 0.3. The beam size is set to 4 for beam search. We initialize the model parameters randomly using a Gaussian distribution with Xavier scheme (Glorot and Bengio, 2010).

We first pre-train the network by minimizing only the maximum likelihood (ML) loss discussed in Eq 13. Next, we initialize our model with the pre-trained ML weights and train the network with the mixed-objective learning function as in Eq 14. The following values of hyperparameters were used to generate the results : (i) $\gamma_1 = 0.99$, $\gamma_2 = 0.01$, $\gamma_3 = 0.1$, (ii) $d_1 = 300$, $d_2 = d_3 = 3$, (iii) $\alpha = 0.9, \beta = 10$, $h = 5000$. Adam (Kingma and Ba, 2014) optimizer is used to train the model with (i) $\beta_1 = 0.9$, (ii) $\beta_2 = 0.999$, and (iii) $\epsilon = 10^{-8}$. For MTL-QG training, the initial learning rate is set to 0.01. For our proposed model training the learning rate is set to 0.00001. We also apply gradient clipping (Pascanu et al., 2013) with range $[-5, 5]$. Based on grid-search, mini-batch size of 16 results in chosen for quick training and fast convergence. To train the supporting facts prediction model, we use the same dataset split as discussed in the Section 4.2 and follow the hyper-parameter setting as discussed in Yang et al. (2018). The optimal beam size =4 is obtained from the model performance on the development dataset.

4.4 Evaluation

We conduct experiments to evaluate the performance of our proposed and other QG methods using the evaluation metrics: BLEU-1, BLEU-2, BLEU-3, BLEU-4 (Papineni et al., 2002), ROUGE-L (Lin, 2004) and METEOR (Banerjee and Lavie, 2005).

Metric for MultiHoping in QG: To assess the multi-hop capability of the question generation model, we introduce additional metric *SF coverage*, which measures in terms of F1 score. This metric is similar

Model	BLEU-4	ROUGE-L	SF Coverage
NQG (Zhou et al., 2017)	19.55	36.63	61.55
SharedEncoder-QG (NQG + Shared Encoder)	20.64	37.18	65.46
MTL-QG (SharedEncoder-QG + SF)	21.28	37.55	70.11
MTL-QG + Rouge-L	22.83	39.41	71.27
Proposed Model (MTL-QG + SF + Rouge-L + MER)	**23.57**	**39.68**	**74.37**

Table 4: A relative performance (on test dataset of HotPotQA) of different variants of the proposed method, by adding one model component.

Model	Naturalness	Difficulty	SF Coverage
NQG	3.20	2.42	73.12
Proposed	**3.47**	**3.21**	**83.04**

Table 5: Human evaluation results for our proposed approach and the NQG model. Naturalness and difficulty are rated on a 1–5 scale and SF coverage is in percentage (%).

to MultiHop-Enhanced Reward, where we use the question-aware supporting facts predictions network that takes the generated question and document list as input and predict the supporting facts. F1 score measures the average overlap between the predicted and ground-truth supporting facts as computed in (Yang et al., 2018).

5 Results and Analysis

We first describe some variants of our proposed MultiHop-QG model.

1. **SharedEncoder-QG**: This is an extension of the NQG model (Zhou et al., 2017) with shared encoder for QG and answer-aware supporting fact predictions tasks. This model is a variant of our proposed model, where we encode the document list using a two-layer Bi-LSTM which is shared between both the tasks. The input to the shared Bi-LSTM is word and answer encoding as shown in Eq. 1. The decoder is a single-layer LSTM which generates the multi-hop question.

2. **MTL-QG**: This variant is similar to the SharedEncoder-QG, here we introduce another Bi-LSTM layer which takes the question, answer and supporting fact embedding as shown in Eq. 3.

The automatic evaluation scores of our proposed method, baselines, and state-of-the-art single-hop question generation model on the HotPotQA test set are shown in Table 2. We also show the additional results on the development dataset in the **Appendix**. The performance improvements with our proposed model over the baselines and state-of-the-arts are statistically significant[2] as ($p < 0.005$). For the question-aware supporting fact prediction model (c.f. 3.1.4), we obtain the F1 and EM scores of 84.49 and 44.20, respectively, on the HotPotQA development dataset. We can not directly compare the result (21.17 BLEU-4) on the HotPotQA dataset reported in Nema et al. (2019) as their dataset split is different and they only use the ground-truth supporting facts to generate the questions.

We also measure the multi-hopping in terms of SF coverage and reported the results in Table 2 and Table 4. We achieve skyline performance of 80.41 F1 value on the ground-truth questions of the test dataset of HotPotQA.

5.1 Quantitative Analysis

Our results in Table 2 are in agreement with (Sun et al., 2018; Zhao et al., 2018; Zhou et al., 2017), which establish the fact that providing the answer tagging features as input leads to considerable improvement in the QG system's performance. Our *SharedEncoder-QG* model, which is a variant of our proposed MultiHop-QG model outperforms all the baselines state-of-the-art models except *Semantic-Reinforced*. The proposed *MultiHop-QG* model achieves the absolute improvement of 4.02 and 3.18 points compared to *NQG* and *Max-out Pointer* model, respectively, in terms of BLEU-4 metric.

To analyze the contribution of each component of the proposed model, we perform an ablation study reported in Table 4. Our results suggest that providing multitask learning with shared encoder helps the model to improve the QG performance from 19.55 to 20.64 BLEU-4. Introducing the supporting facts information obtained from the answer-aware supporting fact prediction task further improves the

[2]We follow the bootstrap test (Efron and Tibshirani, 1994) using the setup provided by Dror et al. (2018).

QG performance from 20.64 to 21.28 BLEU-4. Joint training of QG with the supporting facts prediction provides stronger supervision for identifying and utilizing the supporting facts information. In other words, by sharing the document encoder between both the tasks, the network encodes better representation (supporting facts aware) of the input document. Such presentation is capable of efficiently filtering out the irrelevant information when processing multiple documents and performing multi-hop reasoning for question generation. Further, the MultiHop-Enhanced Reward (MER) with Rouge reward provides a considerable advancement on automatic evaluation metrics. We also perform the experiment with different beam size and reported the performance of QG model is given in the **Appendix**.

5.2 Qualitative Analysis

We have shown the examples in Table 3, where our proposed reward assists the model to maximize the uses of all the supporting facts to generate better human alike questions. In the first example, Rouge-L reward based model ignores the information '*second czech composer*' from the first supporting fact, whereas our MER reward based proposed model considers that to generate the question. Similarly, in the second example, our model considers the information '*disused station located*' from the supporting fact where the former model ignores it while generating the question. We also compare the questions generated from the *NQG* and our proposed method with the ground-truth questions. These questions with additional generated questions are given in the **Appendix**.

Human Evaluation: For human evaluation, we directly compare the performance of the proposed approach with NQG model. We randomly sample 100 document-question-answer triplets from the test set and ask four professional English speakers to evaluate them. We consider three modalities: *naturalness*, which indicates the grammar and fluency; *difficulty*, which measures the document-question syntactic divergence and the reasoning needed to answer the question, and *SF coverage* similar to the metric discussed in Section 4 except we replace the supporting facts prediction network with a human evaluator and we measure the relative supporting facts coverage compared to the ground-truth supporting facts. measure the relative coverage of supporting facts in the questions with respect to the ground-truth supporting facts. *SF coverage* provides a measure of the extent of supporting facts used for question generation. For the first two modalities, evaluators are asked to rate the performance of the question generator on a 1–5 scale (5 for the best). To estimate the *SF coverage* metric, the evaluators are asked to highlight the supporting facts from the documents based on the generated question. We reported the average scores of all the human evaluator for each criteria in Table 5. The proposed approach is able to generate better questions in terms of *Difficulty*, *Naturalness* and *SF Coverage* when compared to the *NQG* model.

6 Conclusion

In this paper, we have introduced the multi-hop question generation task, which extends the natural language question generation paradigm to multiple document QA. Thereafter, we present a novel reward formulation to improve the multi-hop question generation using reinforcement and multi-task learning frameworks. Our proposed method performs considerably better than the state-of-the-art question generation systems on `HotPotQA` dataset. We also introduce SF Coverage, an evaluation metric to compare the performance of question generation systems based on their capacity to accumulate information from various documents. Overall, we propose a new direction for question generation research with several practical applications. In the future, we will be focusing on to improve the performance of multi-hop question generation without any strong supporting facts supervision.

Acknowledgement

Asif Ekbal gratefully acknowledges the Young Faculty Research Fellowship (YFRF) Award supported by the Visvesvaraya PhD scheme for Electronics and IT, Ministry of Electronics and Information Technology (MeitY), Government of India, and implemented by Digital India Corporation (formerly Media Lab Asia).

References

Akari Asai, Akiko Eriguchi, Kazuma Hashimoto, and Yoshimasa Tsuruoka. 2018. Multilingual extractive reading comprehension by runtime machine translation.

Satanjeev Banerjee and Alon Lavie. 2005. Meteor: An automatic metric for mt evaluation with improved correlation with human judgments. In *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization*, pages 65–72. Association for Computational Linguistics.

Yllias Chali and Sadid A. Hasan. 2015. Towards topic-to-question generation. *Computational Linguistics*, 41(1):1–20, March.

Guanliang Chen, Jie Yang, Claudia Hauff, and Geert-Jan Houben. 2018. Learningq: A large-scale dataset for educational question generation. In *Proceedings of the Twelfth International Conference on Web and Social Media, ICWSM 2018, Stanford, California, USA, June 25-28, 2018.*, pages 481–490.

Rotem Dror, Gili Baumer, Segev Shlomov, and Roi Reichart. 2018. The hitchhiker's guide to testing statistical significance in natural language processing. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1383–1392. Association for Computational Linguistics.

Xinya Du, Junru Shao, and Claire Cardie. 2017. Learning to ask: Neural question generation for reading comprehension. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1342–1352. Association for Computational Linguistics.

Bradley Efron and Robert J Tibshirani. 1994. *An introduction to the bootstrap*. CRC press.

Xavier Glorot and Yoshua Bengio. 2010. Understanding the difficulty of training deep feedforward neural networks. In *In Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS'10). Society for Artificial Intelligence and Statistics*.

Caglar Gulcehre, Sungjin Ahn, Ramesh Nallapati, Bowen Zhou, and Yoshua Bengio. 2016. Pointing the unknown words. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 140–149. Association for Computational Linguistics.

Han Guo, Ramakanth Pasunuru, and Mohit Bansal. 2018. Soft layer-specific multi-task summarization with entailment and question generation. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 687–697, Melbourne, Australia, July. Association for Computational Linguistics.

Deepak Gupta, Surabhi Kumari, Asif Ekbal, and Pushpak Bhattacharyya. 2018a. MMQA: A Multi-domain Multi-lingual Question-Answering Framework for English and Hindi. In Nicoletta Calzolari (Conference chair), Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Koiti Hasida, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Asuncion Moreno, Jan Odijk, Stelios Piperidis, and Takenobu Tokunaga, editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May 7-12, 2018. European Language Resources Association (ELRA).

Deepak Gupta, Pabitra Lenka, Asif Ekbal, and Pushpak Bhattacharyya. 2018b. Uncovering code-mixed challenges: A framework for linguistically driven question generation and neural based question answering. In *Proceedings of the 22nd Conference on Computational Natural Language Learning*, pages 119–130.

Deepak Gupta, Asif Ekbal, and Pushpak Bhattacharyya. 2019a. A deep neural network framework for english hindi question answering. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, 19(2):1–22.

Deepak Gupta, Kaheer Suleman, Mahmoud Adada, Andrew McNamara, and Justin Harris. 2019b. Improving neural question generation using world knowledge. *arXiv preprint arXiv:1909.03716*.

Michael Heilman and Noah A Smith. 2010. Good Question! Statistical Ranking for Question Generation. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 609–617. Association for Computational Linguistics.

Yanghoon Kim, Hwanhee Lee, Joongbo Shin, and Kyomin Jung. 2018. Improving neural question generation using answer separation. *arXiv preprint arXiv:1809.02393*.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. *arXiv preprint arXiv:1412.6980*.

Chin-Yew Lin. 2004. ROUGE: A Package for Automatic Evaluation of Summaries. In *Text Summarization Branches Out: Proceedings of the ACL-04 workshop*, volume 8. Barcelona, Spain.

David Lindberg, Fred Popowich, John Nesbit, and Phil Winne. 2013. Generating natural language questions to support learning on-line. In *Proceedings of the 14th European Workshop on Natural Language Generation*, pages 105–114.

Thang Luong, Hieu Pham, and Christopher D. Manning. 2015. Effective approaches to attention-based neural machine translation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1412–1421. Association for Computational Linguistics.

Sewon Min, Victor Zhong, Richard Socher, and Caiming Xiong. 2018. Efficient and robust question answering from minimal context over documents. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1725–1735, Melbourne, Australia, July. Association for Computational Linguistics.

Preksha Nema, Akash Kumar Mohankumar, Mitesh M. Khapra, Balaji Vasan Srinivasan, and Balaraman Ravindran. 2019. Let's ask again: Refine network for automatic question generation. 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)*, pages 3305–3314, Hong Kong, China, November. Association for Computational Linguistics.

Tri Nguyen, Mir Rosenberg, Xia Song, Jianfeng Gao, Saurabh Tiwary, Rangan Majumder, and Li Deng. 2016. Ms marco: A human generated machine reading comprehension dataset. November.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. BLEU: A Method for Automatic Evaluation of Machine Translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pages 311–318. Association for Computational Linguistics.

Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio. 2013. On the difficulty of training recurrent neural networks. In *Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28*, ICML'13, pages III–1310–III–1318. JMLR.org.

Romain Paulus, Caiming Xiong, and Richard Socher. 2018. A deep reinforced model for abstractive summarization. In *ICLR*. OpenReview.net.

Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. GloVe: Global Vectors for Word Representation. In *Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543.

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*, pages 2383–2392. Association for Computational Linguistics.

Steven J Rennie, Etienne Marcheret, Youssef Mroueh, Jerret Ross, and Vaibhava Goel. 2017. Self-critical sequence training for image captioning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7008–7024.

Abigail See, Peter J. Liu, and Christopher D. Manning. 2017. Get to the point: Summarization with pointer-generator networks. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1073–1083. Association for Computational Linguistics.

Min Joon Seo, Sewon Min, Ali Farhadi, and Hannaneh Hajishirzi. 2017a. Query-reduction networks for question answering. In *ICLR*. OpenReview.net.

Minjoon Seo, Aniruddha Kembhavi, Ali Farhadi, and Hananneh Hajishirzi. 2017b. Bidirectional attention flow for machine comprehension. In *International Conference on Learning Representations*.

Jasdeep Singh, Bryan McCann, Nitish Shirish Keskar, Caiming Xiong, and Richard Socher. 2019. Xlda: Cross-lingual data augmentation for natural language inference and question answering. *arXiv preprint arXiv:1905.11471*.

Nitish Srivastava, Geoffrey E Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of machine learning research*, 15(1):1929–1958.

Xingwu Sun, Jing Liu, Yajuan Lyu, Wei He, Yanjun Ma, and Shi Wang. 2018. Answer-focused and position-aware neural question generation. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3930–3939. Association for Computational Linguistics.

Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to Sequence Learning with Neural Networks. In *Advances in Neural Information Processing Systems*, pages 3104–3112.

Alon Talmor and Jonathan Berant. 2018. The web as a knowledge-base for answering complex questions. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 641–651. Association for Computational Linguistics.

Yi Tay, Luu Anh Tuan, and Siu Cheung Hui. 2018. Multi-range reasoning for machine comprehension. *arXiv preprint arXiv:1803.09074*.

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*, pages 191–200, Vancouver, Canada, August. Association for Computational Linguistics.

Tong Wang, Xingdi Yuan, and Adam Trischler. 2017a. A Joint Model for Question Answering and Question Generation. *CoRR*, abs/1706.01450.

Wenhui Wang, Nan Yang, Furu Wei, Baobao Chang, and Ming Zhou. 2017b. 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.

Dirk Weissenborn, Georg Wiese, and Laura Seiffe. 2017. Making neural QA as simple as possible but not simpler. In *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)*, pages 271–280, Vancouver, Canada, August. Association for Computational Linguistics.

Johannes Welbl, Pontus Stenetorp, and Sebastian Riedel. 2018. Constructing datasets for multi-hop reading comprehension across documents. *Transactions of the Association of Computational Linguistics*, 6:287–302.

Ronald J Williams and David Zipser. 1989. A learning algorithm for continually running fully recurrent neural networks. *Neural computation*, 1(2):270–280.

Ronald J Williams. 1992. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 8(3-4):229–256.

Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, et al. 2016. Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144*.

Zhilin Yang, Peng Qi, Saizheng Zhang, Yoshua Bengio, William Cohen, Ruslan Salakhutdinov, and Christopher D. Manning. 2018. Hotpotqa: A dataset for diverse, explainable multi-hop question answering. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2369–2380. Association for Computational Linguistics.

Shiyue Zhang and Mohit Bansal. 2019. Addressing semantic drift in question generation for semi-supervised question answering. 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)*, pages 2495–2509, Hong Kong, China, November. Association for Computational Linguistics.

Yuyu Zhang, Hanjun Dai, Zornitsa Kozareva, Alexander J Smola, and Le Song. 2018. Variational reasoning for question answering with knowledge graph. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

Yao Zhao, Xiaochuan Ni, Yuanyuan Ding, and Qifa Ke. 2018. Paragraph-level neural question generation with maxout pointer and gated self-attention networks. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3901–3910. Association for Computational Linguistics.

Qingyu Zhou, Nan Yang, Furu Wei, Chuanqi Tan, Hangbo Bao, and Ming Zhou. 2017. Neural question generation from text: A preliminary study. In *National CCF Conference on Natural Language Processing and Chinese Computing*, pages 662–671. Springer.

Xinxin Zhu, Lixiang Li, Jing Liu, Longteng Guo, Zhiwei Fang, Haipeng Peng, and Xinxin Niu. 2018. Image captioning with word gate and adaptive self-critical learning. *Applied Sciences*, 8(6):909.

Model	BLEU-1	BLEU-2	BLEU-3	BLEU-4	METEOR	ROUGE-L	SF Coverage
s2s (Du et al., 2017)	35.00	22.57	16.80	13.29	14.90	29.12	60.31
s2s+copy	33.52	23.47	18.46	15.27	18.81	30.66	60.45
s2s+answer	39.72	27.25	20.82	16.62	17.58	33.94	61.14
NQG (Zhou et al., 2017)	42.67	30.53	23.86	19.42	20.69	36.79	61.68
ASs2s (Kim et al., 2018)	42.18	29.97	23.30	18.87	21.27	37.34	63.98
Max-out Pointer (Zhao et al., 2018)	42.05	30.48	24.29	20.17	20.17	34.93	64.19
Semantic-Reinforced (Zhang and Bansal, 2019)	43.84	32.84.	26.02	21.43	23.47	39.17	68.25
SharedEncoder-QG (Ours)	44.51	31.95	25.04	20.40	21.66	37.83	65.22
MTL-QG (Ours)	44.10	32.39	25.94	21.58	21.76	37.77	69.24
Proposed Model	**46.95**	**38.85**	**27.83**	**23.25**	**22.93**	**39.74**	**73.82**

Table 6: Automatic evaluation scores of the proposed approach and the baselines on the development set of HotPotQA.

Model	BLEU-4	ROUGE-L	METEOR	SF Coverage
Proposed Model	**23.57**	**39.68**	22.88	**73.82**
w/ Beam Size 3	23.29	39.67	22.69	73.47
w/ Beam Size 5	23.47	39.60	**22.98**	**73.82**
w/ Beam Size 7	23.37	39.54	23.13	73.54
w/ Beam Size 10	23.12	39.42	23.18	72.94

Table 7: Performance of question generation on the HotPotQA test dataset by varying beam size.

A Question-Aware Supporting Fact Prediction Network

The network uses a convolutional network to obtain a character-based word embedding, which is concatenated with a pre-trained word embedding, and followed by a recurrent layer to encode the contextual information. The same is applied to both question and document. Thereafter, a bi-directional attention layer (Seo et al., 2017b) is employed to fuse the representations of the question and the documents. After using another recurrent layer, we add a self-attention layer (Wang et al., 2017b) followed by a residual connection. For each candidate sentence in the document list, we concatenate the output of the self-attention layer at the first and last positions, and use a binary linear classifier to predict the probability that the current sentence is a supporting fact.

B Analysis

We compare the questions generated from the *NQG* and our proposed method with the ground-truth questions, examples of which are shown in Table 8. The very first example focuses on both the supporting facts, while the NQG model only uses one supporting fact. In the second example, the NQG system generates the question by considering the target answer and a single document. Unlike, our proposed model uses the relation information (*shinga* is a village in *east mamprusi district*) that exists between the entities across the document. It is because our proposed model focuses on all the supporting facts for question generation. We have also shown the examples in Table 10, where our proposed reward assists the model to maximize the uses of all the supporting facts to generate better human alike questions.

Document (1): (a) the m6 motorway runs from junction 19 of the m1 at the catthorpe interchange , near rugby via birmingham then heads north , passing stoke - on - trent , liverpool , manchester , preston , lancaster , carlisle and terminating at the gretna junction (j45) . **(b)** here , just short of the scottish border it becomes the a74(m) which continues to glasgow as the m74 . **Document (2): (a)** shap is a linear village and civil parish located among fells and isolated dales in eden district , cumbria , england . **(b)** the village lies along the a6 road and the west coast main line , and is near to the m6 motorway . **(c)** it is situated 10 mi from penrith and about 15 mi from kendal , in the historic county of westmorland . **Target Answer:** m6 motorway **Reference:** *what motoway runs from junction 19 of the m1 and is near the linear village shap ?* **NQG:** what motorway runs from junction 19 of the m1 at the catthorpe interchange ? **Proposed:** what motorway runs from junction 19 of the m1 at the catthorpe interchange and is near shap ?
Document (1): (a) the east mamprusi district is one of the twenty (20) districts in the northern region of north ghana . **(b)** the capital is gambaga . **Document (2): (a)** shienga (shinga) is a village in east mamprusi district , of the northern region of ghana . **(b)** it lies at an elevation of 349 meters near the right (southern) bank of the white volta . **Target Answer:** gambaga **Reference:** *what is the capitol of the district that also includes the village of shienga ?* **NQG:** what is the capital of the east mamprusi district ? **Proposed:** what is the capital of the district in which shienga is located ?
Document (1): (a) robert clinton smith (born march 30 , 1941) is an american politician who served as a member of the united states house of representatives for new hampshire 's 1st congressional district from 1985 to 1990 and the state of new hampshire in the united states senate from 1990 to 2003 . **Document (2): (a)** new hampshire 's 1st congressional district covers the southeastern part of new hampshire . **(b)** the district consists of three general areas : greater manchester , the seacoast and the lakes region . **Target Answer:** three **Reference:** *bob smith served as a member of the united states house of representatives for a district that consists of how many general areas ?* **NQG:** bob smith served as a member of how many general areas ? **Proposed:** bob smith served as a member of the united states house of representatives for a district that consists of how many general areas ?
Document (1): (a) leanne rowe (born 1982) is an english actress and singer , known for portraying nancy in " oliver twist " , may moss in " lilies " and baby in " dirty dancing : the classic story on stage " . **Document (2): (a)** oliver twist is a 2005 drama film directed by roman polanski **(b)** the screenplay by ronald harwood is based on the 1838 novel of the same name by charles dickens . **Target Answer:** roman polanski **Reference:** *who directed the 2005 film in which leanne rowe portrayed nancy ?* **NQG:** who directed the film in which leanne rowe played nancy ? **Proposed:** who directed the 2005 drama film in which leanne rowe played nancy ?
Document (1): (a) not without laughter is the debut novel by langston hughes published in 1930 . **Document (2): (a)** james mercer langston hughes (february 1 , 1902 – may 22 , 1967) was an american poet . **(b)** he was a social activist , novelist , playwright , and columnist from joplin , missouri . **Target Answer:** american **Reference:** *what was the nationality of the author of " not without laughter " ?* **NQG:** not without laughter is a novel by a man of what nationality ? **Proposed:** not without laughter is the debut novel by the poet of what nationality ?

Table 8: Samples generated by multi-hop question generation approach. In each document, the supporting facts are shown in blue and the target answer is in red.

Model	BLEU-1	BLEU-2	BLEU-3	BLEU-4	METEOR	ROUGE-L	SF Coverage
NQG (Zhou et al., 2017)	42.67	30.53	23.86	19.42	20.69	36.79	61.68
SharedEncoder-QG (NQG + Shared Encoder)	44.51	31.95	25.04	20.40	21.66	37.83	65.22
MTL-QG (SharedEncoder-QG + SF)	44.10	32.39	25.94	21.58	21.76	37.77	69.24
MTL-QG + BLEU	47.29	33.12	26.14	22.46	22.19	39.93	70.15
MTL-QG + Rouge-L	47.16	34.81	27.72	22.83	22.78	40.04	70.87
Proposed Model (MTL-QG + SF + Rouge-L + MER)	**46.95**	**38.85**	**27.83**	**23.25**	**22.93**	**39.74**	**73.82**

Table 9: A relative performance (on development dataset of HotPotQA) of different variants of the proposed method, by adding one model component.

Document (1): (a) orange is a town in grafton county , new hampshire , united states . the population was 331 at the 2010 census . **Document (2): (a)** cardigan mountain state park is a 5655 acre state park in orange , new hampshire . **(b)** the park is free to use , open year - round , and offers a hiking trail up to the 3,121-foot treeless granite summit of mount cardigan . **(c)** there are picnic facilities . **Target Answer:** 5655 acre **Reference:** *how big is the state park located in grafton county , new hampshire ?* **NQG:** how acres is the state park located in orange , new hampshire ? **with only Rouge-L reward:** how many acres is the park in which orange , new hampshire is located ? **with Rouge-L reward and MER:** how many acres is the state park in which town in grafton county , new hampshire ?
Document (1): (a) buddy stephens is an american football coach who is currently the head coach at east mississippi community college , where he has won three njcaa national championships and coached players such as chad kelly and john franklin iii . **(b)** with an overall record of 87–12 , stephens has a higher winning percentage (.879) than the njcaa all - time leader (butler cc 's troy morrell at 154–22 for .875) , but has not yet coached the required 100 games to appear on the list . **Document (2): (a)** john franklin iii is an american football wide receiver for the florida atlantic owls football . **(b)** he formerly played for florida state university , east mississippi community college and auburn university . **Target Answer:** john franklin iii **Reference:** *what player did buddy stephens coach who went on to play for florida state university , east mississippi community college and auburn university ?* **NQG:** buddy stephens is an american football coach who is currently the head coach at east mississippi community college , where he has won three njcaa national championships and coached players such as chad kelly and which american football wide receiver for the florida atlantic owls football ? **with only Rouge-L reward:** what american football wide receiver formerly played for the florida atlantic owls football wide receiver for the florida atlantic owls football , coached buddy stephens ? **with Rouge-L reward and MER:** what is the name of the chad kelly american football coach who formerly formerly played for florida state university , east mississippi community college and auburn university ?
Document (1): (a) seedley railway station is a disused station located in the seedley area of pendleton , salford , on the liverpool and manchester railway . **(b)** it was opened on 1 may 1882 and closed on 2 january 1956 . **(c)** parts of the station wall can still be seen but part of the trackbed has been covered over following the construction of the m602 motorway . **Document (2): (a)** pendleton is an inner city area of salford in greater manchester , england . **(b)** it is about 2 mi from manchester city centre . **(c)** the a6 dual carriageway skirts the east of the district . **Target Answer:** england **Reference:** *seedley railway station is a disused station located in the seedley area of pendleton , is an inner city area of salford in greater manchester , in which country ?* **NQG:** what country does [UNK] railway station and pendleton , greater manchester have in common ? **with only Rouge-L reward** : seedley railway station is located in a city area of salford in what country ? **with Rouge-L reward and MER:** seedley railway station is a disused station located in a city area of salford in greater manchester , in which country ?

Table 10: Sample questions, where our proposed MER based reward model generating better questions than only Rouge-L reward and NQG model.