Learning to Generate Questions with Adaptive Copying Neural Networks

Xinyuan Lu
Carleton University
Ottawa, Canada
xinyuanlu@cmail.carleton.ca

ABSTRACT

Automatic question generation is an important problem in natural language processing. In this paper, we propose a novel adaptive copying recurrent neural network model to tackle the problem of question generation from sentences and paragraphs. The proposed model adds a copying mechanism component onto a bidirectional LSTM architecture to generate more suitable questions adaptively from the input data. Our experimental results show the proposed model can outperform the state-of-the-art question generation methods in terms of BLEU and ROUGE evaluation scores.

KEYWORDS

question generation; copy mechanism; LSTM

ACM Reference Format:
Xinyuan Lu. 2019. Learning to Generate Questions with Adaptive Copying Neural Networks. In *2019 International Conference on Management of Data (SIGMOD'19), June 30-July 5, 2019, Amsterdam,Netherlands.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3299869.3300100

1 INTRODUCTION

Automatic question generation has recently received increasing attention in the natural language processing (NLP) research community. The task is to generate proper questions from a given a sentence or paragraph, which has many applications in NLP, including generating questions for reading comprehension materials and developing dialog systems for building chat robots [5]. In this paper, we propose a new adaptive copying neural network (ACNN) model to tackle the drawbacks of the previous works and generate proper questions. The proposed model exploits a bidirectional LSTM network with global attention mechanism to encode the sequential semantic information of the input sentence or paragraph. When generating semantic questions with a LSTM decoder, it further incorporates a copying mechanism component to allow more suitable and natural works to be properly generated from the source input sequence in a data adaptive manner.

2 RELATED WORK

Question Generation (QG) has drawn a lot attention in these years. The previous work in [9] applied minimal recursion semantics (MRS) to represent the meaning of sentences and then transfer MRS to questions. The work in [3] proposed an overgenerate-and-rank approach to generate and select high quality questions via ranking. The authors of [1] focused on generating questions based on topics. The work in [4] used ontology crowd-sourcing to encode the original text in an ontology and align the question templates to select the most relevant ones. The work in [8] took the knowledge based information as input and generated questions based on it, while [10]combined the answer position of the text. The recent work in [2] proposed a neural question generation model based on LSTM which demonstrates good empirical results.

3 MODEL

Given an input sentence or paragraph $\mathbf{x}$ which is a sequence of tokens $[x_1, ..., x_N]$, we aim to generate a natural question $\mathbf{y} = [y_1, ..., , y_{|}\mathbf{y}|]$ from it.

3.1 Attention Based Encoder

As in [2], we use a bidirectional LSTM to encode the given sequence of tokens in the input sentence $\mathbf{x}$. Let $\overrightarrow{h_t}$ denote the hidden state at time step t for the forward LSTM and $\overleftarrow{h_t}$ for the backward LSTM. The bidirectional LSTM produces the hidden states by $\overrightarrow{\mathbf{h}_t} = \overrightarrow{LSTM}(x_t, \overrightarrow{\mathbf{h}_{t-1}})$, $\overleftarrow{\mathbf{h}_t} = \overleftarrow{LSTM}(x_t, \overleftarrow{\mathbf{h}_{t+1}})$. By concatenating the hidden states from both directions we have the following context dependent hidden representation at time step t $\mathbf{h}_t = [\overrightarrow{\mathbf{h}_t}; \overleftarrow{\mathbf{h}_t}]$. The attention based encoding of $\mathbf{x}$ at a decoding time step k is then computed as a weighted

SIGMOD '19, June 30-July 5, 2019, Amsterdam, Netherlands

ACM ISBN 978-1-4503-5643-5/19/06.
https://doi.org/10.1145/3299869.3300100

average of the representation vectors across $\mathbf{h}_t$.Then, the attention weights $\{a_{k,t}\}$ are calculated using a softmax normalization where W_h is the model parameter to be learned and d_k is the hidden decoding state at time step k which we will introduce below.

$$\mathbf{c}_k = \sum_{t=1}^{N} a_{k,t}\mathbf{h}_t \tag{1}$$

$$a_{k,t} = \frac{exp(e_{k,t})}{\sum_{j=1}^{N} exp(e_{k,j})} \tag{2}$$

$$e_{k,t} = tanh(\mathbf{d}_k^\top W_h \mathbf{h}_t) \tag{3}$$

3.2 Copy Mechanism Based Decoder

The decoding process is to generate question $\mathbf{y}$ from the given sentence $\mathbf{x}$, which is a probabilistic sequence prediction that can be factorized as:

$$P(\mathbf{y}|\mathbf{x}) = \prod_{k=1}^{|\mathbf{y}|} P(y_k|y < k, \mathbf{x}) \tag{4}$$

This is also the empirical probability we need to maximize in the training process across all the annotated training instances. We compute the local conditional word output probability $P(y_k|y < k, \mathbf{x})$ by integrating both a LSTM attention-based decoding component and a copying mechanism component developed in [2], such that

$$P(y_k|y < k, \mathbf{x}) = z_k P_{cop}(y_k) + (1 - z_k)P_{att}(y_k) \tag{5}$$

The attention part $P_{att}(y_k)$ generates words from the common decoder vocabulary, and it is computed on the attention vector $\mathbf{c}_k$ and the hidden state vector $\mathbf{d}_k$ from a decoding LSTM:

$$P_{att}(y_k) = softmax(W_y tanh(W_k[\mathbf{d}_k; \mathbf{c}_k])) \tag{6}$$

where W_y and W_k are model parameters. The copying mechanism component $P_{cop}(y_k)$ generates (copies) words from the individual vocabulary of the source input sequence. We compute it as

$$P_{cop}(y_k) = softmax(V^\top(V[\mathbf{d}_k; \mathbf{c}_k] + b_1) + b_2) \tag{7}$$

when y_k is from the unique word set of the source input sequence, where V, b_1 and b_2 are model parameters. Such a copying mechanism can help incorporating words from the original data into the generated questions. The combination weight z_t is the switch for deciding generating the word from the vocabulary or copying it from the input sequence. Inspired by [7], we calculate z_t as follows.

$$z_t = \sigma(W_d^\top \mathbf{d}_k + W_c^\top \mathbf{c}_k + W_s^\top y_{t-1} + b) \tag{8}$$

where W_d, W_c, W_s and b are model parameters, and σ denotes a sigmoid function. Hence z_k functions as a selection gate that makes data adaptive selection between the attention component and the copying component.

4 EXPERIMENTS AND CONCLUSION

4.1 Experiment Results

Model	BLEU-1	BLEU-2	BLEU-3	BLEU-4	ROUGE-*L*
Seq2Seq	31.34	13.79	7.36	4.26	29.75
Du-*sent*	**43.09**	**25.96**	**17.5**	**12.28**	**39.75**
Du-*para*	42.54	25.33	16.98	11.86	39.37
ACNN-*sent*	**44.78**	**26.83**	**18.72**	**13.97**	**41.08**
ACNN-*para*	44.37	26.15	18.02	13.49	40.57

Table 1: The comparison results in terms of BLEU and ROUGE scores. The best scores in baselines and ACNN are highlighted using boldface.

The experimental results in terms of BLEU 1-4 and ROUGE-*L* scores for all the comparison methods are reported in Table 1. BLEU-n (Bilingual Evaluation Understudy) is a score with n grams calculating the correspondence between machine generated output and ground truth. ROUGE (Recall-Oriented Understudy for Gisting Evaluation) measures the co-occurrences between the system-generated summary and the content in a human-generated summary. ROUGE-L measures the co-occurrences of the longest common subsequence. The best results among the comparison baselines and the proposed ACNN variants are highlighted using boldface font separately.

We can see that among the comparison methods, *Du-sent* produced the best performance in terms of all the evaluation metrics. *Du-para*, though incorporated paragraph, produced slightly inferior performance. Both *Du-sent* and *Du-para* greatly outperform *Seq2Seq*. Among the two variants of our proposed model, ACNN-*sent* slightly outperforms ACNN-*para* which uses paragraph as inputs. This is consistent with the *Du-* methods and might be caused by the noise in the paragraphs. Nevertheless, both variants of the proposed ACNN model consistently outperform the other comparison methods in terms of all the evaluation metrics. The comparison results between ACNN-*sent* and *Du-sent* validate the effectiveness of incorporating the copying mechanism into the bidirectional LSTM question generation model.

4.2 Conclusion

In this paper, we proposed an adaptive copying neural network (ACNN) model for question generation. We incorporated a copying mechanism component into a bidirectional LSTM model with global attention mechanism to improve its capacity on generating proper natural questions. We conducted experiments on the widely used *SQuAD* dataset and showed the proposed model outperforms the state-of-the-art methods in the literature in terms of two types of evaluation metrics, BLEU-*n* and ROUGE-*L*.

A OVERALL STRUCTURE OF ACNN

Figure 1 is the overall structure of our ACNN network.It includes encoder and decoder parts.

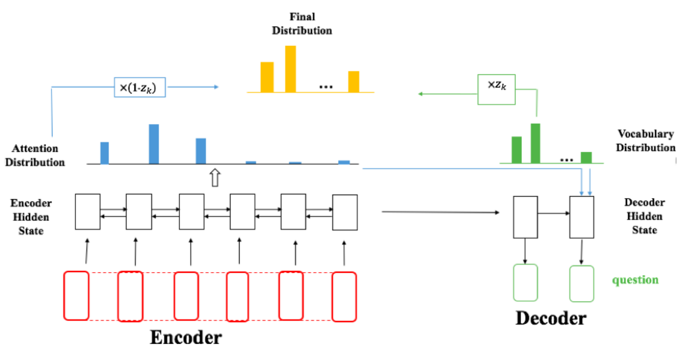

Figure 1: An overall structure of our Adaptive Copying Neural Networks (ACNN)

B DATASET SETTING UP

We conducted experiments on the widely used Stanford Question Answering Dataset (SQuAD)[6]. We used the version released by [2]. It was split into three parts - training set, developing set and test set. The training set contains 70,484 input-question pairs, the development set contains 10,570 input-question pairs, and the test set contains 11,877 input-question pairs.

C IMPACT OF PARAGRAPH LENGTH

Model	BLEU-1	BLEU-2	BLEU-3	BLEU-4	ROUGE-*L*
ACNN-para-150	43.97	25.63	17.48	12.91	39.95
ACNN-para-120	44.22	25.94	17.80	13.26	40.33
ACNN-para-100	**44.37**	**26.15**	**18.02**	**13.49**	**40.57**

Table 2: The results in terms of BLEU and ROUGE scores with different paragraph lengths. The best scores are highlighted using boldface.

We also investigated the impact of the paragraph length on the performance of the proposed variant ACNN-*para*. We tested three different length values, 100, 120 and 150. The comparison results are reported in Table 2. We can see that when increasing the paragraph length, the test scores decrease. This validates our analysis above on paragraph introducing noice: Although longer paragraphs contain more contextual information, they include more irrelevant noisy information as well.

REFERENCES

[1] Yllias Chali and Sadid A. Hasan. 2015. Towards Topic-to-question Generation. *Comput. Linguist.* 41, 1 (March 2015), 1–20. https://doi.org/10.1162/COLI_a_00206
[2] Xinya Du, Junru Shao, and Claire Cardie. 2017. Learning to Ask: Neural Question Generation for Reading Comprehension. *CoRR* abs/1705.00106 (2017). arXiv:1705.00106 http://arxiv.org/abs/1705.00106
[3] Michael Heilman. 2011. *Automatic Factual Question Generation from Text.* Ph.D. Dissertation. Pittsburgh, PA, USA. Advisor(s) Smith, Noah A. AAI3528179.
[4] Igor Labutov, Sumit Basu, and Lucy Vanderwende. 2015. Deep Questions without Deep Understanding. In *Proceedings of ACL*. ACL. https://www.microsoft.com/en-us/research/publication/deep-questions-without-deep-understanding/
[5] Nasrin Mostafazadeh, Ishan Misra, Jacob Devlin, Larry Zitnick, Margaret Mitchell, Xiaodong He, and Lucy Vanderwende. 2016. Generating Natural Questions About an Image. *CoRR* abs/1603.06059 (2016). arXiv:1603.06059 http://arxiv.org/abs/1603.06059
[6] Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. 2016. SQuAD: 100, 000+ Questions for Machine Comprehension of Text. *CoRR* abs/1606.05250 (2016). arXiv:1606.05250 http://arxiv.org/abs/1606.05250
[7] Abigail See, Peter J. Liu, and Christopher D. Manning. 2017. Get To The Point: Summarization with Pointer-Generator Networks. *CoRR* abs/1704.04368 (2017). arXiv:1704.04368 http://arxiv.org/abs/1704.04368
[8] Iulian Vlad Serban, Alberto García-Durán, Çaglar Gülçehre, Sungjin Ahn, Sarath Chandar, Aaron C. Courville, and Yoshua Bengio. 2016. Generating Factoid Questions With Recurrent Neural Networks: The 30M Factoid Question-Answer Corpus. *CoRR* abs/1603.06807 (2016). arXiv:1603.06807 http://arxiv.org/abs/1603.06807
[9] Xuchen Yao. 2010. Question Generation with Minimal Recursion Semantics. (01 2010).
[10] Qingyu Zhou, Nan Yang, Furu Wei, Chuanqi Tan, Hangbo Bao, and Ming Zhou. 2017. Neural Question Generation from Text: A Preliminary Study. *CoRR* abs/1704.01792 (2017). arXiv:1704.01792 http://arxiv.org/abs/1704.01792