IOP Conference Series: Materials Science and Engineering

PAPER • **OPEN ACCESS**

Using Multiple Encoders for Chinese Neural Question Generation from the Knowledge Base

To cite this article: Meixi Chen *et al* 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **490** 042013

Using Multiple Encoders for Chinese Neural Question Generation from the Knowledge Base

Meixi Chen, Jiahao Zhao and Ming Liu*

College of Computer and Information Science, Southwest University, Chongqing, China

*Corresponding author e-mail: mingliu@swu.edu.cn

Abstract. Question generation is an important task in the field of natural language processing and intelligent tutoring system. Previous work on Chinese question generation focused on the rule-based approach, which requires a large amount of human resource to develop the question generation rules. With the recent success of deep neural network in natural language processing, especially the encoder-decoder neural network framework in machine translation, this study explored the effectiveness of the encoder-decoder network in Chinese question generation, where a triple from the knowledge base as an input is encoded and a question as the output is decoded. More importantly, the traditional encoder-decoder network is extended to have multiple encoders that can capture more diverse features to represent the triple. The study results showed that the model with multiple encoders outperformed the traditional encoder-decoder neural network by 1.78 BLEU points.

1. Introduction

Automatic question generation has attracted interest in recent years followed by the increasing interest from the Natural Language Generation (NLG) community [1]. Automatic question generation task has the potential value of question answering (QA) system, reading comprehension [2]. Many prior works mainly applied extraction or transformation rules to generate questions [1]. Mitkov [3] developed a question generation system. This system used rules for creating questions from shallow parses of specific types of sentences. But the study results show that this traditional question generation method do not perform well. Recently, some works using deep neural network to generate questions have made some achievements. Serban et al. [4] adopted a neural network architecture to transform facts into questions. The facts come from the knowledge bases Freebase which consists of a set of triples. Du et al. [5] framed English question generation model for reading comprehension using sequence-to-sequence neural model. Subramanian et al. [6] adopted a sequence-to-sequence model to produce questions based on the key phrases. All the methods proposed above for English question generation adapts the sequence-to-sequence neural model to generate questions and this architecture has made promising achievements on question generation task.

There is a little work about Chinese question generation with the machine learning approach. Motivated by these recent success of deep neural network in English question generation, this paper focuses on Chinese question generation based on deep neural network. This work uses Knowledge Base Question Generation dataset (KBQG) [7] to train Chinese neural question generation model with multiple-encoders. This multiple-encoders framework for Chinese neural question generation is inspired by using multiple encoders and decoders in neural machine translation [8]. In KBQG dataset,

IOP Conf. Series: Materials Science and Engineering **490** (2019) 042013 doi:10.1088/1757-899X/490/4/042013

each instance consists of a triple and a question. A triple is represented by a subject, a relationship and an object, while a question is related to the given triple. The object of the triple is the expected answer of the related question. Some examples of Chinese questions generated by the multiple-encoders model proposed in this paper are shown in Fig. 1. For example, given a triple (Xuanyuan Sword, developers, Softstar Entertainment Inc), the system automatically generates a question asking about who is the publisher of Xuanyuan sword. This question is fluent and similar to the human question (Do you know who is the developer of Xuanyuan sword?).

It can be observed that Chinese neural question generation model with multiple encoders can produce fluent questions and the object of the triple is the expected answer of generated question. The contributions of our work are shown as follows:

1. Previous works [9] for Chinese question generation propose ways to map unstructured text to questions. But Chinese neural question generation model with multiple encoders uses structured text which is represented by a triple to generate questions.
2. Different from Seq2Seq model which is made up of one encoder and one decoder. Chinese neural question generation model with multiple encoders is made up of three encoders and one decoder with attention mechanism. The three encoders separately encode the subject, the relationship and the object from a triple. The model adopts attention mechanism to make multiple-encoders model focus on the three encoders when generating target words.
3. In this study, the questions produced by multiple-encoders model are evaluated by BLEU evaluation metric. In addition, this study also conducts experiments using Seq2Seq model and Seq2Seq attention model. The experiment results show that this multiple-encoders model can get higher BLEU points.

Input triple: （Xuanyuan Sword, developers, Softstar Entertainment Inc） Reference question: Do you know who is the developer of Xuanyuan sword? Output question: I want to know who is the publisher of Xuanyuan sword?
Input triple: （The Legend of Condor Hero, production company, MediaCorp) Reference question: Which company produced The legend of Conder Hero? Output question: Which company produced The Legend of Condor Hero?
Input triple: (La Chapelle, province, Allier) Reference question: What province is La Chapelle in? Output question: Which area does La Chapelle belong to?

Figure 1. Examples of questions generated by multiple-encoders model.

2. Related Work

Question generation task has a lot of potential value for Reading Comprehension, Dialogue System, and Intelligent Assistant System [1]. Tang et al. [10] studied the problem of joint question answering and question generation. Chali and Hasan [11] proposed to apply heuristic rules to transform a sentence into related questions. Heilman and Smith [12] proposed to overproduce questions and rank questions. The strategy improves the question fluency by generating many candidate questions and ranking these questions using a ranking model of question quality. But these methods for automatic question generation do not scale well across different domains [6] and cannot avoid some errors when preprocessing text.

Recently, some question generation strategies which is to apply deep neural network have been proposed. Zhou et al. [13] have proposed to use the neural encoder-decoder model to produce fluency and diverse English question. Their neural question generation framework adds the answer position indicator and lexical features when inputting sentence into the encoder. The study results show that their model can produce more fluent and diverse questions.

Our question generation task shares some similarities with that of Serban et al. [4] which is to generate questions from structured text. Serban et al. [4] adopted a neural Seq2Seq model to generate English questions from triples. This paper uses a multiple-encoders model which consists of three encoders and one decoder with attention mechanism to generate Chinese questions from triples. The three encoders separately process the subject, the relationship, and the object from a triple. And then the decoder with attention mechanism generates a question based on the encoded triple.

3. Chinese Neural Question Generation Model with Multiple Encoders

Our model extends sequence-to-sequence models by using multiple encoders and one decoder with attention mechanism to generate questions focused on answers. Fig. 2 provides an overview of multiple-encoders model. This model separately encodes the subject, the relationship and the object from structured text, and the attention mechanism of the model is based on three encoders. Therefore the decoder can automatically capture important information of a triple that is relevant to predicting target word.

In this study, the proposed multiple-encoders model uses bidirectional LSTM [14] to build each encoder. And this model employs an attention-based LSTM decoder to decode each phrase information which is encoded by the three encoders. In Seq2Seq attention model, the words of the article are fed one-by-one into the encoder, producing a sequence of encoder hidden states $(h_1, h_2, \cdots, h_n)$. Then the model will compute the probability distribution $(a_1, a_2, \cdots, a_n)$ of hidden states. However, multiple-encoders model only compute the probability distribution of the last hidden state $Enc_i(h_{end})$ of each encoder. That is, this multiple-encoders model computes the probability distribution of the three encoders.

3.1. Multiple Encoders

This multiple-encoders model for Chinese question generation uses bidirectional LSTM to build encoders. Each encoder reads different word vector $\mathrm{x} = (x_1, x_2, \cdots, x_n)$ to produce two sequences of hidden states, i.e., the forward sequence $(\overrightarrow{h_1}, \overrightarrow{h_2}, \cdots, \overrightarrow{h_n})$ and the backward sequence $(\overleftarrow{h_1}, \overleftarrow{h_2}, \cdots, \overleftarrow{h_n})$. Next, the forward sequence and the backward sequence are concatenated to form the final hidden states $\mathrm{h} = (h_1, h_2, \cdots, h_n)$:

$$\overrightarrow{h_t} = \overrightarrow{LSTM}(x_t, \overrightarrow{h_{t-1}}) \quad (1)$$

$$\overleftarrow{h_t} = \overleftarrow{LSTM}(x_t, \overleftarrow{h_{t-1}}) \quad (2)$$

$$h_t = [\overrightarrow{h_t}, \overleftarrow{h_t}] \quad (3)$$

The multiple-encoders model builds three encoders which separately read the subject, the relationship and the object of a triple to produce three sequences of hidden states, i.e., $Enc_1(h_1, h_2, \cdots, h_n)$, $Enc_2(h_1, h_2, \cdots, h_n)$ and $Enc_3(h_1, h_2, \cdots, h_n)$.

3.2. Decoder

For the decoder, this model employs a LSTM Networks with attention mechanism on the three encoders. The decoder computes a distribution over three encoders. In more detail, at every time-step t, the decoder computes a soft-alignment score over each last hidden state of three encoders. The attention distribution is calculated as in (4) (5):

$$e_i^t = v^T tanh(W_h Enc_i(h_{end}) + W_s s_{t-1} + b_{attn}) \ \ i = 1,2,3 \quad (4)$$

$$a^t = softmax(e^t) \quad (5)$$

$$s_t = f(s_{t-1}, y_{t-1}, h_t^*) \quad (6)$$

Where v, W_h, W_s, b_{attn} are learnable parameters, s_t represents current decoder state. During decoding, the decoder receives the word embedding of previous word y_{i-1}, the last decoder state s_{t-1} and the current context vector h_t^* to compute current decoder state s_t. $Enc_i(h_{end})$ represents the last hidden state of encoder i. The distribution can be viewed as a probability distribution over the three encoders that tells the decoder which encoder to focus on to produce the next word. Next, the attention distribution is used to produce current context vector h_t^* which is a weighted sum of the last hidden states of the three encoders.

$$h_t^* = \sum_{i=1}^{3} a_i^t Enc_i\,(h_{end}) \ \ i = 1,2,3 \quad (7)$$

The context vector h_t^* can be seen as a fixed-size representation of what has been read from a tripe at current time-step t. Finally, the vocabulary distribution p_{vocab} is computed as in (8):

$$p_{vocab} = softmax(V'(V[s_t, h_t^*] + b) + b') \tag{8}$$

Where V',V,b,b' are learnable parameters. p_{vocab} is the predicted word over all words in vocabulary at time-step t. The loss function of Chinese neural question generation model with multiple-encoders is the negative log-likelihood of the target word w_t^* for time-step t:

$$loss_t = -\log p_{vocab}(w_t^*) \tag{9}$$

The loss for the whole sequence is:

$$\text{loss} = \frac{1}{T}\sum_{t=1}^{T} loss_t \tag{10}$$

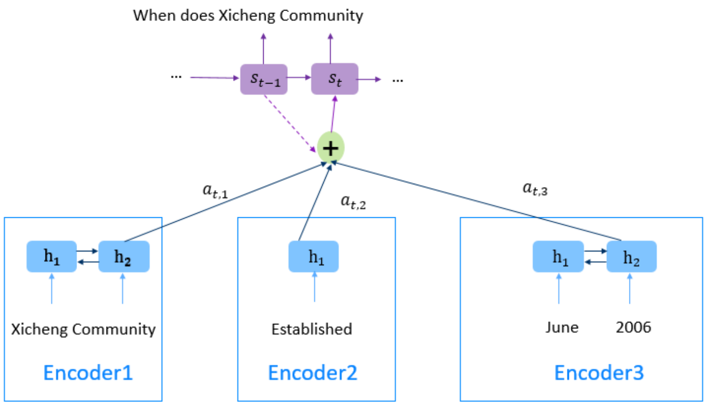

Figure 2. An overview of Chinese question generation model with multiple encoders.

4. Dataset

In this study, we use KBQG dataset [7] to train Chinese neural question generation model with multiple encoders. This dataset is made up of (subject, relationship, object)-question pairs. Every question is related to the given triple represented by a subject, a relationship and an object in KBQG dataset. In detail, the question can be answered by the object entity of the given triple. And both the subject entity and the relationship are explicitly given in each question. For example, in the question (Where is the capital of China?) The question answer is given by the object (Beijing) of the triple (China, capital, Beijing). The subject (China) and the relationship (capital) appear in the question. The statistics of KBQG dataset are shown in Table 1. KBQG dataset is divided into a training set and a test set. There are 23,479 training samples and 1,000 test samples. We use Word Segmentation Tool jieba [15] to segment the text of KBQG dataset and collect the top 30,000 frequent words in training set to build vocabulary for our Chinese multiple-encoders model.

Table 1. Statistics of KBQG dataset.

Dataset	Number of triples	Tokens
Training set	23,479	50k
Test set	1000	10k

5. Experiments and Evaluation

5.1. Experiments

For all experiments, Chinese neural question generation model with multiple encoders has 256 dimensional LSTM hidden states for each encoder and decoder [16]. The word embedding size is set to 300. Parameters of model are updated by Mini-batch Gradient Descent and learning rate is 0.15. According to the statistics of the word length of triples in dataset, we truncate the subject, the relationship and the object to 10 tokens and control the number of words in a question to 30 for training and testing. We use beam search with beam size 6 to generate questions at test time [17]. Multiple-encoders model is trained for about 200,000 iterations and the mini-batch size for the update is set at 32.

We compare our model to two common neural encoder-decoder models. Seq2Seq model [18] is made up of two recurrent neural network (RNN). One RNN encodes a sequence of symbols (subject, relationship, object), while the other RNN decodes the semantic representation of the triple into a sequence of words as a question. Seq2Seq attention model [19] extends Seq2Seq model to encode the input triple into a sequence of vectors and choose a subset of these vectors adaptively to compute current context vector at decode time. Then the model predicts a target word based on the current context vectors associated with these source positions.

We implement all models in Tensorflow. A tripe is directly input into an encoder when training Seq2Seq model and Seq2Seq attention model. The two models also has 256 dimensional BiLSTM hidden states for encoder. The other parameters like training times, the learning rate, vocabulary size are the same with multiple-encoders model for Chinese neural question generation.

5.2. Evaluation
To investigate the performance of three models, we adopt BLEU to measure the quality of questions generated by three models. BLUE is a widely used evaluation metric in machine translation and text summary task. It computes n-gram (n=1, 2, 3, 4) matches between generated and reference questions.

5.3. Results and Discussion
The BLEU1-4 scores of three models are shown in Table 2. It can be observed that multiple-encoders model for Chinese question generation achieves the best performance across BLEU1-4 metrics. This shows that use multiple encoders and make the attention focus on encoders can benefit question generation. This reason is that multiple encoders can provide more comprehensive and accurate representation of a triple and decoder can capture key information at decoding time. Besides, it is noted that Seq2Seq model performs poorly, indicating that this model is not suitable for this task. The Seq2Seq attention model performs better than Seq2Seq model. Because the decoder of Seq2Seq attention model can automatically capture important parts of a source sentence that are relevant to predicting word. Besides, this attention mechanism makes a Seq2Seq model cope with long sentences better.

Table 2. BLEU1-4 evaluation scores of three models.

Model	BLEU-1	BLEU-2	BLEU-3	BLEU-4
Seq2Seq model	25.41	22.69	19.95	19.53
Seq2Seq attention model	26.06	23.85	20.59	20.56
Multiple-encoders model	32.03	23.62	22.67	21.31

6. Conclusion and Future Work
This paper proposed a multiple-encoders model for Chinese question generation which consists of three encoders and one decoder with attention mechanism. This novel architecture can enable the encoders to represent comprehensive and accurate representation of a triple and has a major positive impact on the ability of question generation model to generate good questions. The produced questions are evaluated using BLEU scores. It can be observed that multiple-encoders model outperforms Seq2Seq model and Seq2Seq attention model. But this multiple-encoders model may produce inaccurate interrogative pronouns and not deal with out-of-vocabulary words. In future work, we would like to research how to incorporate pointer network in our model to solve these problems.

Acknowledgments
This work is supported by the National Nature Science Foundation of China(61502397).

References
[1] Heilman M. Automatic Factual Question Generation from Text[C]// Carnegie Mellon University, 2012.
[2] Lopez V, Uren V, Sabou M, et al. Is Question Answering fit for the Semantic Web?: A survey.[J]. Semantic Web, 2011, 2(2):125-155.
[3] Mitkov R. Computer-aided generation of multiple-choice tests[C]// International Conference on Natural Language Processing and Knowledge Engineering, 2003. Proceedings. IEEE, 2003:15.

[4] Serban I V, Garcíadurán A, Gulcehre C, et al. Generating Factoid Questions With Recurrent Neural Networks: The 30M Factoid Question-Answer Corpus[J]. 2016.
[5] Du X, Shao J, Cardie C. Learning to Ask: Neural Question Generation for Reading Comprehension[J]. 2017:1342-1352.
[6] Subramanian S, Wang T, Yuan X, et al. Neural Models for Key Phrase Detection and Question Generation[J]. 2017.
[7] Information on http://tcci.ccf.org.cn/conference/2018/taskdata.php
[8] Zhang J, Liu Q, Zhou J, et al. ME-MD: An Effective Framework for Neural Machine Translation with Multiple Encoders and Decoders[C]// Twenty-Sixth International Joint Conference on Artificial Intelligence. 2017:3392-3398.
[9] Liu M, Rus V, Liu L. Automatic Chinese Factual Question Generation[J]. IEEE Transactions on Learning Technologies, 2017, 10(2):194-204.
[10] Tang D, Duan N, Qin T, et al. Question Answering and Question Generation as Dual Tasks[J]. 2017.
[11] Chali Y, Hasan S A. Towards topic-to-question generation[M]. MIT Press, 2015.
[12] Heilman M, Smith N A. Question Generation via Overgenerating Transformations and Ranking[J]. Question Generation Via Overgenerating Transformations & Ranking, 2009, 8(4):401-407.
[13] Zhou Q, Yang N, Wei F, et al. Neural Question Generation from Text: A Preliminary Study[J]. 2017:662-671.
[14] Graves A, Ndez S, Schmidhuber J, et al. Bidirectional LSTM Networks for Improved Phoneme Classification and Recognition[C]// Artificial Neural Networks: Formal MODELS and Their Applications - ICANN 2005, International Conference, Warsaw, Poland, September 11-15, 2005, Proceedings. DBLP, 2005:799-804.
[15] Information on https://pypi.org/project/jieba/
[16] See A, Liu P J, Manning C D. Get To The Point: Summarization with Pointer-Generator Networks[J]. 2017:1073-1083.
[17] Sammut C. Beam Search[J]. 2011:93-93.
[18] Cho K, Van Merrienboer B, Gulcehre C, et al. Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation[J]. Computer Science, 2014.
[19] Bahdanau D, Cho K, Bengio Y. Neural Machine Translation by Jointly Learning to Align and Translate[J]. Computer Science, 2014.