Reinforced Zero-Shot Cross-Lingual Neural Headline Generation

Ayana, Yun Chen, Cheng Yang, Zhiyuan Liu, and Maosong Sun

***Abstract*—Cross-lingual neural headline generation (CNHG), which aims at training a single, large neural network that directly generates a target language headline given a source language news document, has received considerable attention in recent years. Unlike conventional neural headline generation, CNHG faces the problem that there are no large-scale parallel corpora of source language articles and target language headlines. Consequently, CNHG is a zero-shot scenario. To solve this problem, we propose zero resource CNHG with reinforcement learning. We develop a reinforcement learning framework that is composed of two modules: a neural machine translation (NMT) module and a CNHG module. The translation module translates an input document into a source language document, and the headline generation module takes the previous output as input to generate a target language headline. Then, both modules receive a reward for joint training. The experimental results reveal that our method significantly outperforms baseline models.**

***Index Terms*—Neural networks, cross-lingual headline generation (CNHG), reinforcement learning.**

I. Introduction

THE aim of automatic text summarization is to help individuals handle the problem of data overload. A text summarization model takes one or more documents as input and automatically outputs an informative, coherent, and concise summary. When the length of this summary is shorter than one sentence, this task is called headline generation. Headlines not only play important roles in disseminating information, but also establish connections between readers and the original news documents; that is, readers can decide whether to read the original news documents depending on the headlines.

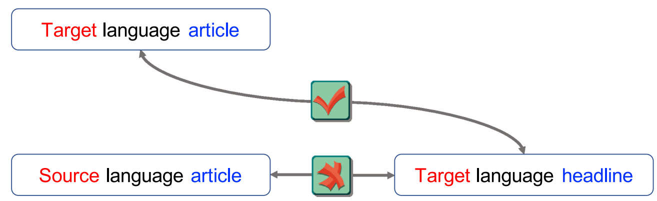

Fig. 1. Presence of training data for NHG models.

Previous studies on automatic headline generation [1]–[5], have required hand-crafted features to measure the informativeness of sentences. Then, various linguistically inspired post-processing is performed to finally create a headline. In contrast, neural headline generation (NHG), benefitting from an end-to-end framework [6], can directly transform the original input article into a corresponding headline. Current NHG models utilize an encoder to read the original input document and encode it into one single fixed-sized vector or a set of fixed-sized vectors from which a decoder generates an output headline word by word. Although the NHG system has many advantages, the training of a NHG system strongly depends on the size of the training data. For example, the Gigaword [7] data required to train an English headline generation system contains nearly 4 million training data pairs, while the LCSTS [8] data required to train a Chinese headline generation system contains nearly 2 million training data pairs. In the absence of a training data, it is difficult to construct and train a NHG model.

Cross-lingual NHG (CNHG) faces such difficulties, as there are no large-scale training data for training the CNHG generation model, as illustrated in Figure 1. A CNHG model aims to read a source language news document and output the corresponding target language headline using a large-scale neural network. The automatic generation of cross-lingual headlines has an important aim: when individuals read the news, they often filter content of interest through the headlines. If the headline is written in a reader's native language, the process of filtering the content of interest becomes more convenient. Ayana *et al.* [9] propose a cross-lingual headline generation system based on the teacher-student framework. They take a pretrained NMT model or a pretrained monolingual NHG model as the teacher models, and generate pseudo headlines using the teacher models. Then the intended CNHG model is trained based on the pseudo target headlines. This framework can directly model the generation process from source language news articles to target language

Manuscript received October 25, 2019; revised March 17, 2020 and June 2, 2020; accepted July 9, 2020. Date of publication July 16, 2020; date of current version September 14, 2020. This work is supported by National Key R&D Program of China (2019AAA0105200) and Beijing Academy of Artificial Intelligence (BAAI). The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Kai Yu. *(Corresponding author: Zhiyuan Liu.)*

Ayana is with the Department of Computer Information Management, Inner Mongolia University of Finance and Economics, Hohhot 010051, China, and also with the Department of Computer Science and Technology State Key Laboratory of Intelligent Technology and Systems Tsinghua, National Laboratory for Information Science and Technology, Tsinghua University, Beijing 100084, China (e-mail: ayana@163.com).

Yun Chen is with the School of Information Management and Engineering, Shanghai University of Finance and Economics, Shanghai 200433, China (e-mail: yunchen@sufe.edu.cn).

Cheng Yang, Zhiyuan Liu, and Maosong Sun are with the Department of Computer Science and Technology State Key Laboratory of Intelligent Technology and Systems Tsinghua, National Laboratory for Information Science and Technology, Tsinghua University, Beijing 100084, China (e-mail: albert.yang33@gmail.com; liuzy@tsinghua.edu.cn; sms@mail.tsinghua.edu.cn).

Digital Object Identifier 10.1109/TASLP.2020.3009487

2329-9290 © 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Tsinghua University. Downloaded on December 16,2020 at 07:18:41 UTC from IEEE Xplore. Restrictions apply.

headlines, thus avoiding the problems of error propagation and model differences caused by a pipeline method.

Although a model based on the teacher-student framework can somewhat alleviate the problem of zero resources in cross-lingual headline generation, two drawbacks remain: 1) the intended CNHG model in the teacher-student framework only approximates the generation probabilities of the teacher model without any ground truth target training data; and 2) the intended CNHG model and teacher model are trained separately and thus unable to cooperate well.

Duan *et al.* [10] attempt to alleviate this problem by extending the work of Ayana *et al.* [9] under the teacher-student framework. They generate pseudo-sources and use true summaries as the simulation target. Although they successfully address the first drawback of Ayana *et al.* [9], the second drawback remained, i.e., the teacher model and the student model were trained separately.

In this work, we attempt to address the drawbacks of teacher-student framework in a different way, and propose a reinforcement learning based framework to model zero-resource CNHG. There are two modules in our framework: a NMT module that translates news documents into the source language, and a CNHG module that generates target language headlines given source language news documents. The two modules work together to complete the training process of cross-lingual headline generation. To obtain improved performance, we establish appropriate reward functions and jointly optimize the two modules.

The main contributions of this paper are as follows:

- We propose a novel reinforced cross-lingual NHG model with no direct training data. In contrast to previous studies that build zero-shot models using a teacher-student framework, we successfully use the reinforcement learning paradigm to encourage collaboration between the NMT module and CNHG module to achieve improved performance.
- We conduct various experiments to verify different aspects of the framework, including the reward design and the training strategies to determine the most appropriate model setup.
- We carry out an ablation study to better understand the collaboration of the NMT and the CNHG modules.
- Extensive experimental results on two bench-mark datasets reveal that our reinforcement learning based models significantly outperform baseline models.

II. Background

A. Neural Headline Generation

Let $\mathbf{x} = \mathrm{x}_1, \ldots, \mathrm{x}_i, \ldots, \mathrm{x}_M$ represent an input document with M words, and let $\mathbf{y} = \mathrm{y}_1, \ldots, \mathrm{y}_j, \ldots, \mathrm{y}_N$ represent a headline with N words. An end-to-end NHG model directly models the generation probability word by word as follows:

$$\log \Pr(\mathbf{y}|\mathbf{x};\boldsymbol{\theta}) = \sum_{j=1}^{N} \log \Pr(\mathrm{y}_j|\mathbf{x}, \mathbf{y}_{<j};\boldsymbol{\theta}), \tag{1}$$

where $\boldsymbol{\theta}$ represents a set of model parameters, and $\mathbf{y}_{<j} = y_1, \ldots, y_{j-1}$ denotes the partially generated headline.

Given a set of training examples $\{\langle \mathbf{x}^{(t)}, \mathbf{y}^{(t)} \rangle\}_{t=1}^{T}$, the standard training objective is to identify a set of model parameters that can maximize the log-likelihood of the training data as follows:

$$\hat{\boldsymbol{\theta}} = \operatorname*{argmax}_{\boldsymbol{\theta}} \{\mathcal{L}(\boldsymbol{\theta})\}, \tag{2}$$

where

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{t=1}^{T} \log \Pr(\mathbf{y}^{(t)}|\mathbf{x}^{(t)};\boldsymbol{\theta}). \tag{3}$$

As mentioned, large-scale training data for NHG only exist for a single language, and no direct training data exist for different languages. Therefore, training a direct cross-lingual headline generation model remains difficult.

B. Reinforcement Learning

Reinforcement learning is an effective way to improve the model performance of neural network based natural language processing tasks, such as dialogue systems [11], paraphrase generation [12], sentiment analysis [13], and machine translation [14], [15]. In the following, we take the reinforced NMT model as an example to introduce the learning process.

A NMT model can be regarded as an agent that interacts with the environment, where the environment comprises generated target words and source context information. Then, under a certain policy (i.e., NMT model parameters), a corresponding action is taken to output the next target word. Finally, the model is updated through the calculated reward. The reward can be calculated according to the reference translation subsequent to obtaining the entire sample. The training goal of reinforcement learning is to maximize the expected reward. When performing an action to obtain the sample, it is almost impossible to enumerate all possible outputs to calculate the expected reward owing to the enormous target vocabulary size. Therefore, in practical applications, a common method is to approximate the expected value by sampling the output target sentence according to the REINFORCE algorithm [16], which is used herein.

III. Model

In this study, we use the reinforcement learning framework to model CNHG in the absence of direct large-scale cross-lingual headline generation training data.

A. Reinforced Model Framework

Given an English headline generation training dataset $\mathcal{D}_{\mathbf{x}_{\mathrm{E}},\mathbf{y}_{\mathrm{E}}} = \{\langle \mathbf{x}_{\mathrm{E}}^{(t)}, \mathbf{y}_{\mathrm{E}}^{(t)} \rangle\}_{t=1}^{T}$ and a Chinese-English translation training dataset $\mathcal{D}_{\mathbf{x}_{\mathrm{CS}},\mathbf{y}_{\mathrm{ES}}} = \{\langle \mathbf{x}_{\mathrm{CS}}^{(n)}, \mathbf{y}_{\mathrm{EC}}^{(n)} \rangle\}_{n=1}^{N}$, we aim to train a direct CNHG model that can generate a cross-lingual English headline $\mathbf{y}_{\mathrm{E}}$ for a Chinese news document. Our framework is composed of two modules: a NMT module $\Pr(\mathbf{y}_{\mathrm{CS}}|\mathbf{x}_{\mathrm{ES}};\hat{\theta}_{\mathrm{NMT}})$ that is responsible for translating an English news article into Chinese, and a CNHG module that generates an English headline given a Chinese news article $\Pr(\mathbf{y}_{\mathrm{E}}|\mathbf{x}_{\mathrm{C}};\hat{\theta}_{\mathrm{CNHG}})$. Here, $\hat{\theta}_{\mathrm{NMT}}$ and $\hat{\theta}_{\mathrm{CNHG}}$ denote the model parameters, and

Authorized licensed use limited to: Tsinghua University. Downloaded on December 16,2020 at 07:18:41 UTC from IEEE Xplore. Restrictions apply.

TABLE I
NOTATIONS

$(\mathbf{x}_{\mathrm{E}}, \mathbf{y}_{\mathrm{E}})$	English article and headline of parallel corpus for English headline generation model
$(\mathbf{x}_{\mathrm{ES}}, \mathbf{y}_{\mathrm{CS}})$	English sentence and Chinese sentence of parallel corpus for English-Chinese translation model
$\hat{\mathbf{x}}_{\mathrm{C}}$	Chinese translation of English article
$\hat{\mathbf{y}}_{\mathrm{E}}$	Model-generated English headline

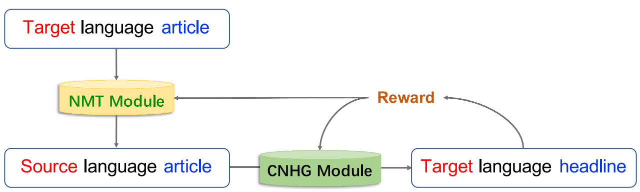

Fig. 2. Overview of proposed model.

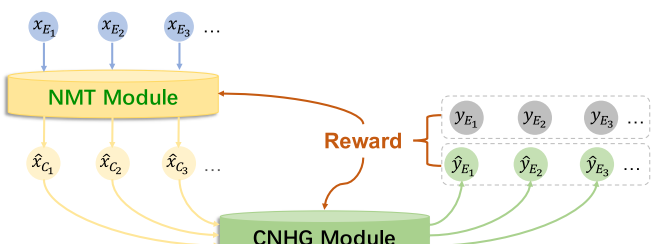

Fig. 3. Overall architecture of reinforced zero-resource headline generation model.

$\Pr(\mathbf{y}_{\mathrm{E}}|\mathbf{x}_{\mathrm{CS}}; \hat{\theta}_{\mathrm{CNHG}})$ is the cross-lingual model that we intend to obtain. For a clear description, the notations are presented in Table I.

As illustrated in Figure 3, the two modules of our framework, the NMT module $\Pr(\mathbf{y}_{\mathrm{E}}|\mathbf{x}_{\mathrm{CS}}; \hat{\theta}_{\mathrm{CNHG}})$ and CNHG module $\Pr(\mathbf{y}_{\mathrm{E}}|\mathbf{x}_{\mathrm{CS}}; \hat{\theta}_{\mathrm{CNHG}})$, not only perform their own tasks but also collaborate, leading to improved performance.

1) The NMT module must correctly translate the English news article $\mathbf{x}_{\mathrm{E}}$ into its corresponding Chinese news article $\hat{\mathbf{x}}_{\mathrm{C}}$, which is the input of the CNHG module;
2) The CNHG module takes the output of the former step as its input and generates a target language headline $\hat{\mathbf{y}}_{\mathrm{E}}$, which then contributes to the calculation of the reward.
3) The reward is calculated based on the target language headline $\hat{\mathbf{y}}_{\mathrm{E}}$ generated in the last step and the reference headline $\mathbf{y}_{\mathrm{E}}$, and the parameters of the translation module and CNHG module are updated based on this reward.

The translation module and cross-lingual headline generation module in this framework can be regarded as two agents, and their parameters can be regarded as the policies in reinforcement learning. For the translation module, the environment is the input English news documents and generated Chinese words, and the action is to generate new translation words according to NMT module parameters. For the cross-lingual generation model, the environment is the Chinese news documents obtained from the translation module and the generated English headline words, and the action is to generate new headline words according to the CNHG module parameters. The reward calculation is performed after the cross-lingual headline generation module generates a complete English headline, and the goal is to maximize the expected value of this reward, as follows:

$$\mathcal{J}_{\mathrm{rl}}(\boldsymbol{\theta}_{\mathrm{NMT}}, \boldsymbol{\theta}_{\mathrm{CNHG}}) = \mathbb{E}_{\hat{\mathbf{x}}_{\mathrm{C}} \sim \Pr(\mathbf{x}_{\mathrm{C}}|\mathbf{x}_{\mathrm{E}}; \boldsymbol{\theta}_{\mathrm{NMT}})} \times \mathcal{R}\left(\hat{\mathbf{x}}_{\mathrm{C}}, \mathbf{y}_{\mathrm{E}}, \boldsymbol{\theta}_{\mathrm{CNHG}}\right). \tag{4}$$

The final training objective is to identify two sets of parameters that can maximize the expected reward as follows:

$$\hat{\theta}_{\mathrm{NMT}}, \hat{\theta}_{\mathrm{CNHG}} = \underset{\boldsymbol{\theta}_{\mathrm{NMT}}, \boldsymbol{\theta}_{\mathrm{CNHG}}}{\operatorname{argmax}} \left\{\mathcal{J}_{\mathrm{rl}}(\boldsymbol{\theta}_{\mathrm{NMT}}, \boldsymbol{\theta}_{\mathrm{CNHG}})\right\}. \tag{5}$$

Setting an appropriate reward function has a large impact on the performance of the entire enhanced learning framework. In this study, it is important to set a proper value for the reward function in (4), or $\mathcal{R}(\hat{\mathbf{x}}_{\mathrm{C}}, \mathbf{y}_{\mathrm{E}}, \boldsymbol{\theta}_{\mathrm{CNHG}})$. We use two methods to define the reward function and investigate their impact on model performance. The reward function is usually defined as non-differentiable discrete values by comparing the generated output and reference output; thus, we use the REINFORCE algorithm [16] to calculate the gradient and update the model parameters. Next, we introduce two reward functions and their corresponding gradient calculation methods.

B. *Reward* 1*: Generation Probability of CNHG Model*

Suppose that a Chinese news document $\hat{\mathbf{x}}_{\mathrm{C}}$ that is sampled from the translation model, and an English reference headline $\mathbf{y}_{\mathrm{E}}$ constitute a training data pair for the CNHG module. The generation probability of the CNHG can be taken as the reward function:

$$\mathcal{R}_1\left(\hat{\mathbf{x}}_{\mathrm{C}}, \mathbf{y}_{\mathrm{E}}, \boldsymbol{\theta}_{\mathrm{CNHG}}\right) = \log \Pr(\mathbf{y}_{\mathrm{E}}|\hat{\mathbf{x}}_{\mathrm{C}}; \boldsymbol{\theta}_{\mathrm{CNHG}}). \tag{6}$$

He *et al.* [17] and Chen *et al.* [18] utilize the generation probability to define the reward function; thus, this study also adopts this method. Given $\mathcal{R}_1$ as a reward function, (4) can be rewritten as follows:

$$\mathcal{J}_{\mathcal{R}_1}(\boldsymbol{\theta}_{\mathrm{NMT}}, \boldsymbol{\theta}_{\mathrm{CNHG}}) = \mathbb{E}_{\hat{\mathbf{x}}_{\mathrm{C}} \sim \Pr(\mathbf{x}_{\mathrm{C}}|\mathbf{x}_{\mathrm{E}}; \boldsymbol{\theta}_{\mathrm{NMT}})} \times \log \Pr(\mathbf{y}_{\mathrm{E}}|\hat{\mathbf{x}}_{\mathrm{C}}; \boldsymbol{\theta}_{\mathrm{CNHG}}). \tag{7}$$

The partial derivative of $\boldsymbol{\theta}_{\mathrm{NMT}}$ and $\boldsymbol{\theta}_{\mathrm{CNHG}}$ can be calculated as:

$$\begin{aligned} &\nabla_{\boldsymbol{\theta}_{\mathrm{NMT}}} \mathcal{J}_{\mathcal{R}_1}(\boldsymbol{\theta}_{\mathrm{NMT}}, \boldsymbol{\theta}_{\mathrm{CNHG}}) \\ &= \mathbb{E}_{\hat{\mathbf{x}}_{\mathrm{C}} \sim \Pr(\mathbf{x}_{\mathrm{C}}|\mathbf{x}_{\mathrm{E}}; \boldsymbol{\theta}_{\mathrm{NMT}})} \left[\mathcal{R}_1 \nabla_{\boldsymbol{\theta}_{\mathrm{NMT}}} \log \Pr\left(\hat{\mathbf{x}}_{\mathrm{C}}|\mathbf{x}_{\mathrm{E}}; \boldsymbol{\theta}_{\mathrm{NMT}}\right)\right] \\ &\nabla_{\boldsymbol{\theta}_{\mathrm{CNHG}}} \mathcal{J}_{\mathcal{R}_1}(\boldsymbol{\theta}_{\mathrm{NMT}}, \boldsymbol{\theta}_{\mathrm{CNHG}}) \\ &= \mathbb{E}_{\hat{\mathbf{x}}_{\mathrm{C}} \sim \Pr(\mathbf{x}_{\mathrm{C}}|\mathbf{x}_{\mathrm{E}}; \boldsymbol{\theta}_{\mathrm{NMT}})} \left[\nabla_{\boldsymbol{\theta}_{\mathrm{CNHG}}} \log \Pr\left(\mathbf{y}_{\mathrm{E}}|\hat{\mathbf{x}}_{\mathrm{C}}; \boldsymbol{\theta}_{\mathrm{CNHG}}\right)\right]. \end{aligned} \tag{8}$$

Due to the enormous vocabulary size, it is intractable to first enumerate all possible candidate translations corresponding to $\mathbf{x}_{\mathrm{E}}$ and then calculate the partial derivatives in (8) because the size of the candidate translation set is exponential. Common practice is to sample one sentence $\hat{\mathbf{x}}_{\mathrm{C}}$ from the entire sample space $\Pr(\mathbf{x}_{\mathrm{C}}|\mathbf{x}_{\mathrm{E}}; \boldsymbol{\theta}_{\mathrm{NMT}})$ to approximate the partial derivative

Authorized licensed use limited to: Tsinghua University. Downloaded on December 16,2020 at 07:18:41 UTC from IEEE Xplore. Restrictions apply.

as follows:

$$\begin{aligned}&\nabla_{\boldsymbol{\theta}_{\mathrm{NMT}}}\mathcal{J}_{\mathcal{R}_1}(\boldsymbol{\theta}_{\mathrm{NMT}},\boldsymbol{\theta}_{\mathrm{CNHG}})\\&\qquad\approx\mathcal{R}_1\nabla_{\boldsymbol{\theta}_{\mathrm{NMT}}}\log\Pr(\hat{\mathbf{x}}_{\mathrm{C}}|\mathbf{x}_{\mathrm{E}};\boldsymbol{\theta}_{\mathrm{NMT}})\\&\nabla_{\boldsymbol{\theta}_{\mathrm{CNHG}}}\mathcal{J}_{\mathcal{R}_1}(\boldsymbol{\theta}_{\mathrm{NMT}},\boldsymbol{\theta}_{\mathrm{CNHG}})\\&\qquad\approx\nabla_{\boldsymbol{\theta}_{\mathrm{CNHG}}}\log\Pr(\mathbf{y}_{\mathrm{E}}|\hat{\mathbf{x}}_{\mathrm{C}};\boldsymbol{\theta}_{\mathrm{CNHG}})\,.\end{aligned}\tag{9}$$

C. *Reward 2: Expected ROUGE Score of CNHG Model*

In reinforcement learning, it is common to use the evaluation metric during testing as the optimization target of the end-to-end model, and it has achieved favorable results in many end-to-end tasks [19]–[22]. ROUGE is a benchmark for headline generation tasks; thus, we use the ROUGE value as a reward function as follows:

$$\begin{aligned}\mathcal{R}_2(\hat{\mathbf{x}}_{\mathrm{C}},\mathbf{y}_{\mathrm{E}},\boldsymbol{\theta}_{\mathrm{CNHG}})&=\mathbb{E}_{\hat{\mathbf{y}}_{\mathrm{E}}\sim\Pr(\mathbf{y}_{\mathrm{E}}|\hat{\mathbf{x}}_{\mathrm{C}};\boldsymbol{\theta}_{\mathrm{CNHG}})}\\&\qquad\times\mathrm{ROUGE}(\hat{\mathbf{y}}_{\mathrm{E}},\mathbf{y}_{\mathrm{E}}).\end{aligned}\tag{10}$$

Given $\mathcal{R}_2$ as the reward function, (4) can be rewritten as:

$$\begin{aligned}&\mathcal{J}_{\mathcal{R}_2}(\boldsymbol{\theta}_{\mathrm{NMT}},\boldsymbol{\theta}_{\mathrm{CNHG}})\\&=\mathbb{E}_{\hat{\mathbf{x}}_{\mathrm{C}}\sim\Pr(\mathbf{x}_{\mathrm{C}}|\mathbf{x}_{\mathrm{E}};\boldsymbol{\theta}_{\mathrm{NMT}})}\left[\mathbb{E}_{\hat{\mathbf{y}}_{\mathrm{E}}\sim\Pr(\mathbf{y}_{\mathrm{E}}|\hat{\mathbf{x}}_{\mathrm{C}};\boldsymbol{\theta}_{\mathrm{CNHG}})}\right.\\&\qquad\left.\times\ \mathrm{ROUGE}(\hat{\mathbf{y}}_{\mathrm{E}},\mathbf{y}_{\mathrm{E}})\right].\end{aligned}\tag{11}$$

The partial derivative with respect to parameters $\boldsymbol{\theta}_{\mathrm{NMT}}$ and $\boldsymbol{\theta}_{\mathrm{CNHG}}$ can be calculated as:

$$\begin{aligned}&\nabla_{\boldsymbol{\theta}_{\mathrm{NMT}}}\mathcal{J}_{\mathcal{R}_2}(\boldsymbol{\theta}_{\mathrm{NMT}},\boldsymbol{\theta}_{\mathrm{CNHG}})\\&=\mathbb{E}_{\hat{\mathbf{x}}_{\mathrm{C}}\sim\Pr(\mathbf{x}_{\mathrm{C}}|\mathbf{x}_{\mathrm{E}};\boldsymbol{\theta}_{\mathrm{NMT}})}\left[\mathcal{R}_2\nabla_{\boldsymbol{\theta}_{\mathrm{NMT}}}\log\Pr(\hat{\mathbf{x}}_{\mathrm{C}}|\mathbf{x}_{\mathrm{E}};\boldsymbol{\theta}_{\mathrm{NMT}})\right]\\&\nabla_{\boldsymbol{\theta}_{\mathrm{CNHG}}}\mathcal{J}_{\mathcal{R}_2}(\boldsymbol{\theta}_{\mathrm{NMT}},\boldsymbol{\theta}_{\mathrm{CNHG}})\\&=\mathbb{E}_{\hat{\mathbf{x}}_{\mathrm{C}}\sim\Pr(\mathbf{x}_{\mathrm{C}}|\mathbf{x}_{\mathrm{E}};\boldsymbol{\theta}_{\mathrm{NMT}})}\Big[\\&\qquad\mathbb{E}_{\hat{\mathbf{y}}_{\mathrm{E}}\sim\Pr(\mathbf{y}_{\mathrm{E}}|\hat{\mathbf{x}}_{\mathrm{C}};\boldsymbol{\theta}_{\mathrm{CNHG}})}\left[\nabla_{\boldsymbol{\theta}_{\mathrm{CNHG}}}\log\Pr(\hat{\mathbf{y}}_{\mathrm{E}}|\hat{\mathbf{x}}_{\mathrm{C}};\boldsymbol{\theta}_{\mathrm{CNHG}})\right.\\&\qquad\qquad\left.\times\mathrm{ROUGE}(\hat{\mathbf{y}}_{\mathrm{E}},\mathbf{y}_{\mathrm{E}})\right]\Big].\end{aligned}\tag{12}$$

Since enumerating all candidates $\hat{\mathbf{x}}_{\mathrm{C}}\sim\Pr(\mathbf{x}_{\mathrm{C}}|\mathbf{x}_{\mathrm{E}};\boldsymbol{\theta}_{\mathrm{NMT}})$ is infeasible as the target vocabulary size and sentence length result in an exponential sample space. The same problem occurs when sampling $\hat{\mathbf{y}}_{\mathrm{E}}\sim\Pr(\mathbf{y}_{\mathrm{E}}|\hat{\mathbf{x}}_{\mathrm{C}};\boldsymbol{\theta}_{\mathrm{CNHG}})$. For $\hat{\mathbf{x}}_{\mathrm{C}}\sim\Pr(\mathbf{x}_{\mathrm{C}}|\mathbf{x}_{\mathrm{E}};\boldsymbol{\theta}_{\mathrm{NMT}})$, we use the same approximation method as in the previous section. For $\hat{\mathbf{y}}_{\mathrm{E}}\sim\Pr(\mathbf{y}_{\mathrm{E}}|\hat{\mathbf{x}}_{\mathrm{C}};\boldsymbol{\theta}_{\mathrm{CNHG}})$, we utilize the same method as in Ayana *et al.* [23]. We randomly sample K_1 samples to approximate the sample space and calculate the ROUGE score for each sample in the sample space to approximate the reward value. Then, the approximated gradient becomes:

$$\begin{aligned}&\nabla_{\boldsymbol{\theta}_{\mathrm{NMT}}}\mathcal{J}_{\mathcal{R}_2}(\boldsymbol{\theta}_{\mathrm{NMT}},\boldsymbol{\theta}_{\mathrm{CNHG}})\\&\qquad\approx\mathcal{R}_2\nabla_{\boldsymbol{\theta}_{\mathrm{NMT}}}\log\Pr(\hat{\mathbf{x}}_{\mathrm{C}}|\mathbf{x}_{\mathrm{E}};\boldsymbol{\theta}_{\mathrm{NMT}})\\&\nabla_{\boldsymbol{\theta}_{\mathrm{CNHG}}}\mathcal{J}_{\mathcal{R}_2}(\boldsymbol{\theta}_{\mathrm{NMT}},\boldsymbol{\theta}_{\mathrm{CNHG}})\\&\qquad\approx\frac{1}{K_1}\sum_{k=1}^{K_1}\Big[\nabla_{\boldsymbol{\theta}_{\mathrm{CNHG}}}\log\Pr\left(\hat{\mathbf{y}}_{\mathrm{E}}^{(k)}|\hat{\mathbf{x}}_{\mathrm{C}};\boldsymbol{\theta}_{\mathrm{CNHG}}\right)\\&\qquad\qquad\times\mathrm{ROUGE}(\hat{\mathbf{y}}_{\mathrm{E}}^{(\mathrm{k})},\mathbf{y}_{\mathrm{E}}).\end{aligned}\tag{13}$$

When using ROUGE to evaluate the performance of the headline generation model, the F values of ROUGE-1, ROUGE-2, and ROUGE-L are generally used for evaluation. ROUGE-1 and ROUGE-2 are used to evaluate the ability of a system to capture salient information, while ROUGE-L is used to evaluate the fluency of the system output. According to a previous study [23], optimizing models using different ROUGE values has different effects on model performance; thus, it is necessary to experimentally identify the most suitable ROUGE value for this study.

D. *Training*

Previous studies [14], [15], [24]–[26] have determined that the training process is unstable if the modules of a reinforcement learning framework are jointly trained from the beginning. We also observed this problem in our training process. In this study, there are two modules: an English-Chinese NMT module that is responsible for translating an English news article $\mathbf{x}_{\mathrm{E}}$ into its corresponding Chinese news article $\hat{\mathbf{x}}_{\mathrm{C}}$, and the intended Chinese-English CNHG module that takes the Chinese news article $\hat{\mathbf{x}}_{\mathrm{C}}$ as input and outputs an English headline $\mathbf{y}_{\mathrm{E}}$. The target vocabulary of the two modules includes tens of thousands of words, and the length of $\hat{\mathbf{x}}_{\mathrm{C}}$ and $\mathbf{y}_{\mathrm{E}}$ reaches 20∼50 and 5∼20 words, respectively. To solve the consequent problem of the enormous and intractable search space of $\hat{\mathbf{x}}_{\mathrm{C}}$ and $\mathbf{y}_{\mathrm{E}}$, we also adopt a sampling method. If we randomly initialized the two modules and used only the reward function to optimize them, the performance of each module would be unstable and unable to converge. To stabilize the training process better, we propose three training strategies, each corresponding to a model variant. We introduce these strategies in greater detail in the following subsections.

Pretraining: We utilize English-Chinese translation training data $\mathcal{D}_{\mathbf{x}_{\mathrm{ES}},\mathbf{y}_{\mathrm{CS}}}=\{\langle\mathbf{x}_{\mathrm{ES}}^{(n)},\mathbf{y}_{\mathrm{CS}}^{(n)}\rangle\}_{n=1}^{N}$ to pretrain an English-Chinese NMT model. The training objective is to maximize the log-likelihood of the training data as follows:

$$\begin{aligned}\hat{\theta}_{\mathrm{NMT_{pre}}}=&\ \underset{\boldsymbol{\theta}_{\mathrm{NMT}}}{\mathrm{argmax}}\\&\times\left\{\sum_{(\mathbf{x}_{\mathrm{ES}},\mathbf{y}_{\mathrm{CS}})\in\mathcal{D}_{x_{\mathrm{ES}},y_{\mathrm{CS}}}}\log\Pr(\mathbf{y}_{\mathrm{CS}}|\mathbf{x}_{\mathrm{ES}};\boldsymbol{\theta}_{\mathrm{NMT}})\right\}\end{aligned}\tag{14}$$

We then use the pretrained model parameter $\hat{\theta}_{\mathrm{NMT_{pre}}}$ to initialize the NMT modules in the reinforcement learning framework. The parameter of the CNHG module is randomly initialized. To prevent the randomly initialized CNHG module from influencing the performance of the NMT module during training, we fix the parameters of the NMT module in the first three training epochs, and only the parameters of the CNHG module are updated. The training process is illustrated in Algorithm 1.

Reinforced training: Subsequent to the training process presented in algorithm 1, we obtain a pretrained CNHG model,

Authorized licensed use limited to: Tsinghua University. Downloaded on December 16,2020 at 07:18:41 UTC from IEEE Xplore. Restrictions apply.

Algorithm 1: Pretraining Process.

Input :
English NHG training data $\mathcal{D}_{\mathbf{x}_{\mathrm{E}},\mathbf{y}_{\mathrm{E}}} = \{\langle \mathbf{x}_{\mathrm{E}}^{(m)}, \mathbf{y}_{\mathrm{E}}^{(m)} \rangle\}_{m=1}^{M}$;
English-Chinese NMT module with parameter of $\hat{\theta}_{\mathrm{NMT}_{\mathrm{pre}}}$;
Learning rate γ_{CNHG};
Output:
Optimized Chinese-English CNHG module parameter $\hat{\theta}_{\mathrm{CNHG}_{\mathrm{pre}}}$;
Randomly initialize Chinese-English NHG module parameter θ_{CNHG};
epoch $\leftarrow 1$;
while *epoch* < 4 **do**
  Receive a random data pair $\langle \mathbf{x}_{\mathrm{E}}, \mathbf{y}_{\mathrm{E}} \rangle$ from training dataset $\mathcal{D}_{\mathbf{x}_{\mathrm{E}},\mathbf{y}_{\mathrm{E}}}$;
  Take $\mathbf{x}_{\mathrm{E}}$ as input, generate $\hat{\mathbf{x}}_{\mathrm{C}}$ using the NMT module with parameter of $\Pr(\hat{\theta}_{\mathrm{NMT}_{\mathrm{pre}}})$;
  if $Reward\ 1$ **then**
    Compute $\mathcal{R}_1(\hat{\mathbf{x}}_{\mathrm{C}}, \mathbf{y}_{\mathrm{E}}, \theta_{\mathrm{CNHG}}) = \log \Pr(\mathbf{y}_{\mathrm{E}} | \hat{\mathbf{x}}_{\mathrm{C}}; \theta_{\mathrm{CNHG}})$;
    Compute the gradient of the objective function w.r.t. θ_{CNHG}: $\nabla_{\theta_{\mathrm{CNHG}}} = \nabla_{\theta_{\mathrm{CNHG}}} \log \Pr(\mathbf{y}_{\mathrm{E}} | \hat{\mathbf{x}}_{\mathrm{C}}; \theta_{\mathrm{CNHG}})$;
  end
  else
    Take $\hat{\mathbf{x}}_{\mathrm{C}}$ as input, generate K_1 English headlines from the current CNHG module;
    Compute $\mathcal{R}_2(\hat{\mathbf{x}}_{C}, \mathbf{y}_{E}, \theta_{\mathrm{CNHG}}) \approx \frac{1}{K_1} \sum_{k=1}^{K_1} \left[\nabla_{\theta_{\mathrm{CNHG}}} \log \Pr\left(\hat{\mathbf{y}}_{\mathrm{E}}^{(k)} | \hat{\mathbf{x}}_{\mathrm{C}}; \theta_{\mathrm{CNHG}} \right) \times \mathrm{ROUGE}(\hat{\mathbf{y}}_{\mathrm{E}}^{(\mathrm{k})}, \mathbf{y}_{\mathrm{E}}) \right]$;
    Compute the gradient of the objective function w.r.t. θ_{CNHG}:
    $\nabla_{\theta_{\mathrm{CNHG}}} = \frac{1}{K_1} \sum_{k=1}^{K_1} \left[\nabla_{\theta_{\mathrm{CNHG}}} \log \Pr\left(\hat{\mathbf{y}}_{\mathrm{E}}^{(k)} | \hat{\mathbf{x}}_{\mathrm{C}}; \theta_{\mathrm{CNHG}} \right) \times \mathrm{ROUGE}(\hat{\mathbf{y}}_{\mathrm{E}}^{(\mathrm{k})}, \mathbf{y}_{\mathrm{E}}) \right]$;
  end
  Update module parameter: $\theta_{\mathrm{CNHG}} \leftarrow \theta_{\mathrm{CNHG}} + \gamma_{\mathrm{CNHG}} \nabla_{\theta_{\mathrm{CNHG}}}$
end

Algorithm 2: Reinforced Training Process.

Input :
English NHG training data $\mathcal{D}_{\mathbf{x}_{\mathrm{E}},\mathbf{y}_{\mathrm{E}}} = \{\langle \mathbf{x}_{\mathrm{E}}^{(m)}, \mathbf{y}_{\mathrm{E}}^{(m)} \rangle\}_{m=1}^{M}$
English-Chinese NMT module with parameter $\hat{\theta}_{\mathrm{NMT}_{\mathrm{pre}}}$
Chinese-English CNHG module with parameter $\hat{\theta}_{\mathrm{CNHG}_{\mathrm{pre}}}$
Learning rate γ_{NMT},γ_{CNHG}
Output:
Optimized English-Chinese NHG module parameter $\hat{\theta}_{\mathrm{NMT}_{\mathrm{rl}}}$
Optimized Chinese-English CNHG module parameter $\hat{\theta}_{\mathrm{CNHG}_{\mathrm{rl}}}$
while *Not Converged* **do**
  Receive a random data pair $\langle \mathbf{x}_{\mathrm{E}}, \mathbf{y}_{\mathrm{E}} \rangle$ from training dataset $\mathcal{D}_{\mathbf{x}_{\mathrm{E}},\mathbf{y}_{\mathrm{E}}}$;
  Take $\mathbf{x}_{\mathrm{E}}$ as input, generate $\hat{\mathbf{x}}_{\mathrm{C}}$ using the NMT module with parameter of $\Pr(\hat{\theta}_{\mathrm{NMT}_{\mathrm{pre}}})$;
  if $Reward\ 1$ **then**
    Compute $\mathcal{R}_1(\hat{\mathbf{x}}_{\mathrm{C}}, \mathbf{y}_{\mathrm{E}}, \theta_{\mathrm{CNHG}}) = \log \Pr(\mathbf{y}_{\mathrm{E}} | \hat{\mathbf{x}}_{\mathrm{C}}; \theta_{\mathrm{CNHG}})$;
    Compute the gradient of the objective function w.r.t. θ_{NMT}: $\nabla_{\theta_{\mathrm{NMT}}} = \mathcal{R}_1 \nabla_{\theta_{\mathrm{NMT}}} \log \Pr(\hat{\mathbf{x}}_{\mathrm{C}} | \mathbf{x}_{\mathrm{E}}; \theta_{\mathrm{NMT}})$
    Compute the gradient of the objective function w.r.t. θ_{CNHG}: $\nabla_{\theta_{\mathrm{CNHG}}} = \nabla_{\theta_{\mathrm{CNHG}}} \log \Pr(\mathbf{y}_{\mathrm{E}} | \hat{\mathbf{x}}_{\mathrm{C}}; \theta_{\mathrm{CNHG}})$;
  end
  else
    Take $\hat{\mathbf{x}}_{\mathrm{C}}$ as input, generate K_1 English headlines from the current CNHG module;
    Compute $\mathcal{R}_2(\hat{\mathbf{x}}_{C}, \mathbf{y}_{E}, \theta_{\mathrm{CNHG}}) \approx \frac{1}{K_1} \sum_{k=1}^{K_1} \left[\nabla_{\theta_{\mathrm{CNHG}}} \log \Pr\left(\hat{\mathbf{y}}_{\mathrm{E}}^{(k)} | \hat{\mathbf{x}}_{\mathrm{C}}; \theta_{\mathrm{CNHG}} \right) \times \mathrm{ROUGE}(\hat{\mathbf{y}}_{\mathrm{E}}^{(\mathrm{k})}, \mathbf{y}_{\mathrm{E}}) \right]$;
    Compute the gradient of the objective function w.r.t. θ_{NMT}: $\nabla_{\theta_{\mathrm{NMT}}} = \mathcal{R}_2 \nabla_{\theta_{\mathrm{NMT}}} \log \Pr(\hat{\mathbf{x}}_{\mathrm{C}} | \mathbf{x}_{\mathrm{E}}; \theta_{\mathrm{NMT}})$
    Compute the gradient of the objective function w.r.t. θ_{CNHG}:
    $\nabla_{\theta_{\mathrm{CNHG}}} = \frac{1}{K_1} \sum_{k=1}^{K_1} \left[\nabla_{\theta_{\mathrm{CNHG}}} \log \Pr\left(\hat{\mathbf{y}}_{\mathrm{E}}^{(k)} | \hat{\mathbf{x}}_{\mathrm{C}}; \theta_{\mathrm{CNHG}} \right) \times \mathrm{ROUGE}(\hat{\mathbf{y}}_{\mathrm{E}}^{(\mathrm{k})}, \mathbf{y}_{\mathrm{E}}) \right]$;
  end
  Update module parameter: $\theta_{\mathrm{NMT}} \leftarrow \theta_{\mathrm{NMT}} + \gamma_{\mathrm{NMT}} \nabla_{\theta_{\mathrm{NMT}}}$
  Update module parameter: $\theta_{\mathrm{CNHG}} \leftarrow \theta_{\mathrm{CNHG}} + \gamma_{\mathrm{CNHG}} \nabla_{\theta_{\mathrm{CNHG}}}$
end

which is parameterized by $\hat{\theta}_{\mathrm{CNHG}_{\mathrm{pre}}}$. In our reinforced training model, we also initialize the NMT module by $\hat{\theta}_{\mathrm{NMT}_{\mathrm{pre}}}$. However, the CNHG module parameter is initialized by pretrained $\hat{\theta}_{\mathrm{CNHG}_{\mathrm{pre}}}$. During the training process, the parameters of the NMT and CNHG modules are updated until the CNHG module converges.

The training process is presented in algorithm 2. The differences between Algorithms 2 and 1 are as follows: 1) the inputs are different, as algorithm 2 also takes a pretrained CNHG model as input; 2) the outputs are different, as algorithm 2 also outputs the optimized NMT parameter; and 3) the calculation steps are different, as algorithm 2 adds the steps of calculating partial

Authorized licensed use limited to: Tsinghua University. Downloaded on December 16,2020 at 07:18:41 UTC from IEEE Xplore. Restrictions apply.

derivatives of the NMT module and updating the parameters of the NMT module. Here, since the NMT module is updated according to the reward value from the intended CNHG module, the NMT module is more compatible with the headline generation module.

Joint training with translation dataset: In the two models discussed above, only English headline generation training data $\mathcal{D}_{x_{\mathrm{E}},y_{\mathrm{E}}}$ are used for training. To constrain the NMT module by its own translation task during the training process, we also use English-Chinese translation training data $\mathcal{D}_{x_{\mathrm{ES}},y_{\mathrm{CS}}}$ to jointly train the NMT module. Specifically, there are two ways of performing the joint training: the first is to maximize the maximum likelihood of the translation data, while the second is to use BLEU as a reward to maximize its expected value. The objective function corresponding to the first method is as follows:

$$\begin{aligned}&\mathcal{J}_{\mathrm{joint}_1}(\boldsymbol{\theta}_{\mathrm{NMT}},\boldsymbol{\theta}_{\mathrm{CNHG}})\\&=\mathbb{E}_{\hat{\mathbf{x}}_{\mathrm{C}}\sim\Pr(\mathbf{x}_{\mathrm{C}}|\mathbf{x}_{\mathrm{E}};\boldsymbol{\theta}_{\mathrm{NMT}})}\mathcal{R}\left(\hat{\mathbf{x}}_{\mathrm{C}},\mathbf{y}_{\mathrm{E}},\boldsymbol{\theta}_{\mathrm{CNHG}}\right)\\&\quad+\lambda_1\sum_{(\mathbf{x}_{\mathrm{ES}},\mathbf{y}_{\mathrm{CS}})\in\mathcal{D}_{x_{\mathrm{ES}},y_{\mathrm{CS}}}}\log\Pr(\mathbf{y}_{\mathrm{CS}}|\mathbf{x}_{\mathrm{ES}};\boldsymbol{\theta}_{\mathrm{NMT}}).\end{aligned}\tag{15}$$

The objective function corresponding to the second method is as follows:

$$\begin{aligned}&\mathcal{J}_{\mathrm{joint}_2}(\boldsymbol{\theta}_{\mathrm{NMT}},\boldsymbol{\theta}_{\mathrm{CNHG}})\\&=\mathbb{E}_{\hat{\mathbf{x}}_{\mathrm{C}}\sim\Pr(\mathbf{x}_{\mathrm{C}}|\mathbf{x}_{\mathrm{E}};\boldsymbol{\theta}_{\mathrm{NMT}})}\mathcal{R}\left(\hat{\mathbf{x}}_{\mathrm{C}},\mathbf{y}_{\mathrm{E}},\boldsymbol{\theta}_{\mathrm{CNHG}}\right)\\&\quad+\lambda_2\mathbb{E}_{\hat{\mathbf{y}}_{\mathrm{CS}}\sim\Pr(\mathbf{y}_{\mathrm{CS}}|\mathbf{x}_{\mathrm{ES}};\boldsymbol{\theta}_{\mathrm{NMT}})}\mathrm{BLEU}(\hat{\mathbf{y}}_{\mathrm{CS}},\mathbf{y}_{\mathrm{CS}}).\end{aligned}\tag{16}$$

During training, we use the method similar to that in Shen *et al.* [19] to approximate the expected reward. Specifically, we sample K_2 translations to approximate the entire sample space to calculate the expected reward.

IV. Experiments

We evaluate our method on a Chinese-English cross-lingual headline generation task. In the following, we first introduce the experimental settings, including the model settings and experimental datasets. We then present the baseline systems for comparison. Finally, we present the experimental results, including the effects of different sampling methods, the effects of different ROUGE scores, and the main experimental results.

A. Setup

NMT model setup: For the English-Chinese NMT module, we use a bidirectional GRU-RNN encoder and an attention-based GRU-RNN decoder as the module architecture. The English vocabulary size is set to 40,000, the Chinese vocabulary size is set to 60,000, and both the word embedding size and hidden size are set to 1,024. During training, sentences containing more than 50 words are removed, the batch size is set to 40, and adadelta [27] is adopted for optimization. We use the open-source NMT toolkit dl4mt[1] implemented by Theano [28] for all experiments. We evaluate the model performance for every 5,000 iterations, and stop the training process when the model performance no longer increase more than 10 times. The translation training data includes 1.25 M pairs of sentences [2] with 34.5 M English words and 27.9 M Chinese words. We use the NIST 2002 dataset as the development set and THULAC[3] as the Chinese word segmentation toolkit. The evaluation metric is BLEU [29], and the running script is *multi-bleu.perl*.

Target CNHG model setup: For the Chinese-English CNHG module, we adopt the same settings as in the NMT module. It should be noted that the Chinese vocabulary of the CNHG module is the same as that of the NMT module because the two modules had to interact with each other in Chinese during reinforced learning. However, because English in the NMT module is in the form of a news document and English in the CNHG module is in the form of a headline, different vocabularies have to be used. The training data is extracted from English Gigaword [7] and contains a total of 3.8 M English document-headline data pairs. ROUGE [30] is used for evaluation, and the corresponding script is *ROUGE-1.5.5.pl*. This script reports the recall, precision, and F value. Because the recall value is easily affected by the generation length, a longer headline obtains a higher recall value; therefore, we use a fairer F value as the evaluation method. The relevant settings for the development set and test set are described in the following section.

Data construction: CNHG is a task that lacks large-scale training data, and thus there were no universal development or test data by the time we stated this work. Therefore, we decide to manually construct Chinese document-English headline development and test sets. For traditional monolingual NHG task, previous studies usually tune their model parameters on DUC2003 dataset and compare model performances on DUC2004 dataset and Gigaword test set. To build the development and test data for the Chinese-English CNHG model, we manually translate the English documents into Chinese and pair the translations with the originally provided English headlines. One professional translator is invited to accomplish the translation process and the following translation instructions are provided to improve the accuracy of the translation process,: 1) the translation should correspond to the source input documents; 2) Since the translated document is a news document, the translation should avoid being overly colloquial; and 3) leave the proper nouns in the original document unchanged to avoid unnecessary ambiguity. Table II presents detailed data statistics.

Shortly afterward, Duan *et al.* [10] also propose to solve the CNHG problem and publicly release their manually built Chinese-English Gigaword test data. To make the experimental results more convincing, we adopt DUC2003 data to tune our model parameters, DUC2004 data and Gigaword test data to test and compare baseline systems.

[1] [Online]. Available: https://github.com/nyu-dl/dl4mt-tutorial

[2] The training data included LDC2002E18, LDC2003E07, LDC2003E14, and part of LDC2004T07, LDC2004T08, and LDC2005T06.

[3] [Online]. Available: http://thulac.thunlp.org/

Authorized licensed use limited to: Tsinghua University. Downloaded on December 16,2020 at 07:18:41 UTC from IEEE Xplore. Restrictions apply.

TABLE II
DATA STATISTICS OF DUC DATASETS. ART.NUM, ART.AVG.TOK, AND HEAD.AVG.TOK REFER TO THE ARTICLE NUMBERS, AVERAGE TOKEN NUMBERS IN EACH TRANSLATED ARTICLE, AND AVERAGE TOKEN NUMBERS IN EACH HEADLINE, RESPECTIVELY

Dataset	Statistics		
	art.num	art.avg.tok	head.avg.tok
DUC2003	624	34.0	10.03
DUC2004	500	35.0	10.43

B. Baseline Systems

In this section, we introduce the baseline systems.

- **Baseline-TS** (**T**ranslate first, then **S**ummarize): This is a trivial solution to the zero-resource CNHG problem in a pipeline way. With this method, the original input Chinese news documents are first translated into their corresponding English translations by a pretrained Chinese-English NMT model. Then, the translations are summarized using a pretrained English headline generation model.
- **Baseline-ST** (**S**ummarize first, then **T**ranslate): This method is also a pipeline method. Rather than translating first, as in the previous baseline method, this method utilizes a pretrained Chinese NHG model to generate headlines for the original input news documents, and then translates them into their corresponding English headlines using the pretrained Chinese-English NMT model.
- **Baseline-Teach**: This indicates the teacher-student framework of Ayana *et al.* [9]. This method is built under the assumption that the intended English headline of a Chinese news document would have the same generation probability with the translation of Chinese headline and the headline of English news document. Hence, the student model is the intended Chinese-English CNHG model and the teacher models are the Chinese-English NMT model and the English NHG. Given a Chinese news document-headline data pair from LCSTS, the student model takes the Chinese news document as the input, and mimic the Chinese-English NMT teacher model translation probability of the Chinese headline and the English NHG teacher model headline generation probability of the pseudo English headline.
- **Baseline-Generation+Attention**: Duan *et al.* [10] also propose to solve the Chinese-English CNHG problem under the teacher-student framework.[4] The main difference between their work and Ayana *et al.* [9] is that they adopt pseudo source input, while Ayana *et al.* [9] adopt pseudo target output. Given the pseudo Chinese input, this method not only uses the cross-entropy to encourage the similarity between generation probabilities of the Chinese-English CNHG student model and the English NHG teacher model, but also utilizes an attention relay mechanism to encourage the consistency of the attention weights between the teacher model and the student model.

Ayana *et al.* [9] and Duan *et al.* [10] also compare their work with a baseline called Baseline-PSEUDO. With this method, they compose a pseudo-training dataset to train the CNHG model. Specifically, they use an English-Chinese NMT model to translate the English document part $\mathbf{x_E}$ from the English document-headline training dataset $\mathcal{D}_{\mathbf{x_E},\mathbf{y_E}}$ using greedy decoding to obtain pseudo-Chinese news document $\hat{\mathbf{x}}_\mathbf{C}$. They then take $\hat{\mathbf{x}}_\mathbf{C}$ and the English headline part $\mathbf{y_E}$ from $\mathcal{D}_{\mathbf{x_E},\mathbf{y}^\mathbf{E}}$ to compose a pseudo-dataset $\hat{\mathcal{D}}_{\hat{\mathbf{x}}_\mathbf{E},\mathbf{y}_C}$ to directly train an end-to-end CNHG model. In our study the pretraining method has the same methodology as Baseline-PSEUDO; that is, it generates pseudo-training data using greedy training. Therefore, we do not report the results specifically, as they correspond to the pretraining model results.

[4]Duan *et al.* [10] adopt Transformer [31] as their basic model architecture, and we utilize the attention-based bi-directional GRU as ours. Hence, for a fair comparison, we do not directly refer to the experimental results from their original paper. Specifically, we re-implement their three model variants with the attention-based bi-directional GRU architecture and report the corresponding experimental results.

C. Effect of Different Sampling Methods

In our reinforcement learning framework, the NMT module is required to take an English news document $\mathbf{x}_\mathrm{E}$ as input and translate it into a Chinese news document $\mathbf{x}_\mathrm{C}$, which is then used as the input of the CNHG module. Thereafter, model training is performed, and the model parameters are updated. It is infeasible to enumerate the entire search space. As a result, we utilize approximation methods to address this problem. In consideration of the time and space complexity, we use the following two methods.

- **Random sampling**: At each decoding step of the NMT module, the decoder takes as input the encoder hidden states, decoder output word from the previous step, and hidden state from the previous step. It then calculates and selects one word as the output through a multinomial distribution. This method not only creates more data diversity but also increases the robustness of the model [32].
- **Greedy sampling**: Although the random sampling method can create data diversity and is less time-consuming, it has a drawback. When a randomly sampled word at each decoding step is less semantically related to the original input, the error propagation problem is more severe. A simple way to overcome this problem is to utilize the greedy decoding method, as suggested by Kim and Rush [33]. Specifically, when the NMT module decoder outputs a word, it always selects the word with the highest generation probability based on the softmax distribution. In this way, the sampled translation is more stable.

We conduct an experiment on the DUC2003 development dataset to verify the performance of the two sampling methods, and the method with superior performance is adopted as the default sampling method in subsequent experiments. To perform a fair comparison, we set the reward function to $Reward\ 1$ and utilized the pretraining method for training. Table III presents the experimental results. It can be seen that the greedy sampling method outperformed random sampling on all three ROUGE evaluation metrics. One possible explanation is that in our

Authorized licensed use limited to: Tsinghua University. Downloaded on December 16,2020 at 07:18:41 UTC from IEEE Xplore. Restrictions apply.

TABLE III
EFFECT OF DIFFERENT SAMPLING METHODS ON THE DUC2003 DEVELOPMENT DATASET. $R1$, $R2$, AND RL REPRESENT THE F SCORE OF ROUGE-1, ROUGE-2, AND ROUGE-L, RESPECTIVELY

Sampling method	Evaluation method		
	$R1$	$R2$	RL
Random sampling	14.40	3.06	13.07
Greedy sampling	**16.00**	**3.26**	**14.20**

TABLE IV
EFFECT OF DIFFERENT ROUGE SCORES ON DUC2003 DEVELOPMENT DATASET

ROUGE	Evaluation method		
	$R1$	$R2$	RL
ROUGE-1	18.79	3.01	16.12
ROUGE-2	15.81	**3.34**	14.24
ROUGE-L	**18.91**	3.06	**16.87**

reinforcement learning framework, we introduce several approximations to increase the efficiency of the training process. In this setting, stability is more important than data diversity. Therefore, we adopt greedy sampling as the sampling method in the following experiments.

D. *Effect of Different ROUGE Scores*

When the reward function is set as $Reward\ 2$, the evaluation metrics of the CNHG module as the reward function of the reinforcement learning framework. During testing, previous studies commonly reported ROUGE-1, ROUGE-2, and ROUGE-L scores to measure system performance from different perspectives. ROUGE-1 and ROUGE-2 focus more on informativeness, while ROUGE-L focuses more on coherence. We set $Reward\ 2$ as the above-mentioned evaluation metrics to investigate the effects of the metrics on system performance. It should also be noted that several previous studies reported recall-based ROUGE scores, while others reported F1-based ROUGE scores. The recall-based scores are easily affected by the length of the system output; that is, the longer the system output, the higher the recall score. The F1-based scores, in contrast, impose corresponding penalties with regard to the length, and provide fairer results. Therefore, we adopted F1-based scores to define $Reward\ 2$.

Table IV presents the experimental results on the DUC2003 development dataset. We determine that when the reward function is defined by ROUGE-L, the CNHG model obtain the highest ROUGE-1 and ROUGE-L scores, while when the reward function is defined by ROUGE-2, the model obtain the highest ROUGE-2 value. Since the model obtain the highest score on two evaluation metrics when the reward function is defined by ROUGE-L, we decide to define the $Reward\ 2$ function using the F1-based ROUGE-L in subsequent experiments.

E. *Main Results*

Table V provides the experimental results on the DUC2004 test dataset. The first row in the table presents the evaluation results of the baseline systems. The second row presents the experimental results when $Reward\ 1$ is adopted as the reward function, that is, the generation probability of the CNHG module when different training strategies are used. The final row corresponds to the experimental results when we utilize $Reward\ 2$ as the reward function and adopt different training strategies. *pretraining*, *reinforced training*, and *joint training* are the training methods that are introduced in Section III-D.

Comparisons between the Baselines: The Baseline-TS and Baseline-ST are two pipeline methods. However, baseline-ST performs significantly worse than baseline-TS. The reason for this is that the first step of the pipeline method (i.e., the summarization step) is completed using the Chinese NHG model, which is trained on the LCSTS data, and the data distribution of LCSTS data is much different from English evaluation data.

The different data distribution problem is also one of the reasons why the directly trained Baseline-Teach [9] model based on the teacher-student framework performs even worse than baseline-ST. One of their teacher networks would be the Chinese NHG model, which is pretrained on the LCSTS dataset. However, the LCSTS dataset is collected from Sina Weibo,[5] and the data distribution differed from the Giga- word and DUC data. As a result, the pseudo-input document generated by the Chinese NHG teacher network would be unsatisfactory, as would the subsequent target distribution generated by the Chinese-English NMT. Besides the data distribution problem, the pseudo-target-side training data would also impair the model performance.

Baseline-Generation+Attention [10] solves the problems of different data distribution and the pseudo-target-side training data under the teacher-student framework. As a result, Baseline-Generation+Attention surpasses all the other three pipeline systems indicating the importance of the ground truth target training data.

Comparison with baseline systems: Comparing the first row of Table V with the second and third rows, our results consistently outperform the two pipeline methods on ROUGE-1 and ROUGE-L scores. We believe that this is because our work models the CNHG task in a direct way, as opposed to the pipeline methods of the baseline models. However, the ROUGE-2 scores of the two pretraining models and the $Reward\ 1$ + reinforced training model are inferior to the baseline-TS model. One possible explanation is that in our framework, when sampling the input $\hat{\mathbf{x}}_C$ for the CNHG module, we only utilize the greedy sampling method to reduce the time complexity. In the baseline-TS model, in contrast, the translations are generated using beam search with BEAM-SIZE $= 10$. This problem can be somewhat alleviated by introducing the translation training data to jointly train the two modules, as our other models outperform the baseline models based on the ROUGE-2 score. Our models also consistently outperform Baseline-Teach on all three evaluation metrics due to the training data distribution problem.

Baseline-Generation+Attention [10] is also a teacher-student framework and shares the same motivation with ours, which is to solve the pseudo target training data problem by constructing a framework that takes the pseudo source data while keeping

[5][Online]. Available: http://www.weibo.com

Authorized licensed use limited to: Tsinghua University. Downloaded on December 16,2020 at 07:18:41 UTC from IEEE Xplore. Restrictions apply.

TABLE V
EXPERIMENTAL RESULTS ON DUC2004 AND GIGAWORD TEST DATASETS

Model		DUC2004			Gigaword		
		$R1$	$R2$	RL	$R1$	$R2$	RL
Baseline-TS		15.95	3.49	14.15	17.98	4.75	16.32
Baseline-ST		12.89	2.29	11.44	13.55	2.44	12.14
Baseline-Teach [9]		12.73	2.02	11.35	12.86	2.44	11.80
Baseline-Generation+Attention [10]		19.49	**4.69**	17.38	20.11	5.94	18.42
$Reward\ 1$	Pretraining	16.50	2.72	14.30	17.75	4.98	16.18
	Reinforced training	17.28	3.31	15.16	17.99	4.99	16.38
	Joint training	18.70	4.14	16.33	19.40	**5.96**	17.91
$Reward\ 2$	Pretraining	20.05	3.15	17.37	21.82	4.43	19.74
	Reinforced training	20.50	3.87	17.90	20.25	5.97	18.47
	Joint training	**20.62**	4.16	**18.02**	**22.35**	5.64	**20.29**

the target data true. Baseline-Generation+Attention not only achieves the best performance comparing to the three other baseline systems but also performs better than some of our models. However, our $Reward\ 2$ + Joint training model performs better than it on ROUGE-1 and ROUGE-L scores indicating that with appropriate reward function and training strategy, our model could achieve more promising results.

Effect of different reward functions: A comparison of the results in the second and third rows of Table V demonstrates that the ROUGE-1 and ROUGE-L scores of the model with $Reward\ 2$ are generally higher than those of the model with $Reward\ 1$. In addition, the score of $Reward\ 2$ + pretraining model is higher than that of $Reward\ 1$ + joint training model. This demonstrates that it is effective to directly adopt the evaluation metric as a reward function in the reinforcement learning framework.

Effect of reinforced training: A comparison between the reinforced training model and pretraining model in each row of Table V reveals that the reinforced training model outperforms the pretraining model on three evaluation metrics. This is because the NMT module and CNHG module of our reinforcement learning framework are jointly trained and updated during training. As a result, the NMT module is more suitable for the preceding part of the cross-lingual headline generation task, provides more appropriate input to the subsequent target CNHG module, and improves the performance of the target CNHG module.

Effect of joint training: We update and train the two modules constituting the entire reinforcement learning framework in the reinforced training model. Because the reward function is calculated based on feedback from the CNHG module, it is natural to update the CNHG module using this reward. However, for NMT modules, there is a concern that training only through feedback from the CNHG module results in the problem that the NMT module lacks constraints from the translation tasks. As a result, we add extra translation training data for joint training on the basis of enhanced training. A comparison of the results of joint training and reinforced training indicates that the evaluation results of the joint training model outperform the reinforced training model in each evaluation metric, which further confirms the effectiveness of joint training. In addition, this method also contributes to improving the ROUGE-2 score of the model, which compensates for the deficiency of the sampling process.

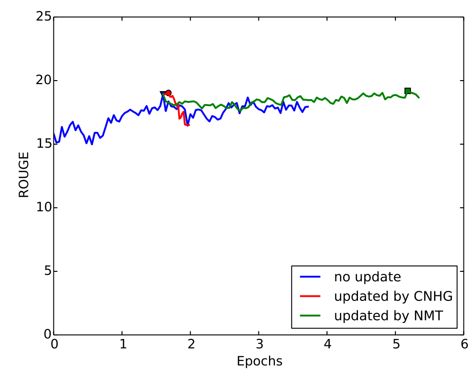

Fig. 4. ROUGE-1 score curves on DUC2003.

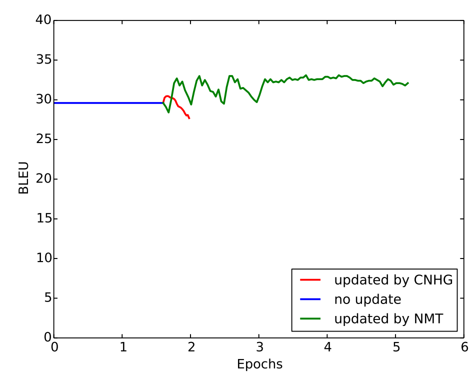

Fig. 5. BLEU score curves on NIST02.

Effect of collaboration: To better understand whether the collaboration between the NMT model and CNHG model is the source of performance improvement, we conduct an ablation study that keeps the reinforcement learning framework but sets different updating strategies to the NMT model during training. We set the reward function to $Reward\ 2$ to perform a fair comparison. Figure 4 shows the ROUGE-1 score curves of the CNHG model and Figure 5 shows the BLEU score curves of the NMT model corresponding to each ablation. We use the DUC2003 dataset as the validation set for CNHG model and NIST02 dataset as the validation set for NMT model.

Authorized licensed use limited to: Tsinghua University. Downloaded on December 16,2020 at 07:18:41 UTC from IEEE Xplore. Restrictions apply.

TABLE VI
SAMPLE HEADLINES FROM BASELINE SYSTEMS AND OUR PROPOSED SYSTEMS

English article		Cambodian politicians expressed hope on Monday that a new partnership between the parties of strongman Hun Sen and his rival, Prince Norodom Ranariddh, in a coalition government would not end in more violence.
Chinese article		柬埔寨政界人士周一 表示，希望铁腕人物洪森和他的对手诺罗敦拉那烈亲王在联合政府中的新伙伴关系不会以更多的暴力结束。
English reference 1		Hun Sen and Prince Ranariddh share power in new coalition Cambodian government.
English reference 2		Cambodian coalition formed between Hun Sen and Ranariddh. Rainsy left out.
English reference 3		Hope for partnership between Hun Sen and rival, Prince Ranariddh.
English reference 4		Cambodians hope that violence will be avoided in new coalition government.
Baseline-ST	Step-1	柬埔寨高官：希望UNK在选举中的新伙伴关系不会更多的暴力结束 (Cambodian senior officer: I hope UNK’s new partnership in the election will not end more violently)
	Step-2	Cambodian senior officials have expressed their wish to see whether the new Zealand - China relations will be more or less violent during the election.
Baseline-TS	Step-1	Cambodian people ’s political circles said on Monday that he hoped the new partnership between the two parties will not end more than ever more violent in his rival in the coalition government.
	Step-2	Cambodian people’s political party says it hopes new alliance will not end violence in coalition government.
$Reward\ 1$	Pretrain	Cambodia’s prince and prince set to form new government.
	Reinforced	Cambodia’s new PM Hun Sen to form new government.
	Joint	Cambodia’s new prime minister Hun Sen to form new government.
$Reward\ 2$	Pretrain	Cambodian government Khmer Rouge to of talks.
	Reinforced	Cambodian Prince Ranariddh to for new partnership with Cambodia.
	Joint	Cambodian political parties concerned over coalition government.

During the pretraining, according to 1, we train the CNHG model for 3 epochs while freezing the NMT model and obtain a set of optimal model parameters, which is then used to initialize the reinforced training model and joint training model. When we keep training the CNHG model under the reinforcement learning framework with the frozen NMT model for two more epochs, we find that the performance of the CNHG model is no longer improved. We then update the NMT model by the CNHG model reward during training. We observe that the CNHG model performance continues to decline after a short period of improvement. The bleu score curve in Figure 5 also shows the same pattern. When we update the NMT model also using the translation training data, the CNHG model performs more steady and generally better than the prior two methods. The corresponding bleu scores in Figure 5 are also higher than the prior two methods. According to the previous observations, we can draw a conclusion that the collaboration of the NMT model and the CNHG model could indeed boost the CNHG model performance.

F. Case Study

Table VI presents headline examples from the baseline systems and our proposed methods. For a fair comparison, we randomly select the input article. To improve readability, we also perform post-editing.

Baseline-ST and baseline-TS are pipeline models. Thus, the results corresponding to each step are listed in the table. In baseline-ST, Step-1 is the Chinese headline generated by the Chinese NHG model, while Step-2 is the English headline generated by the Chinese-English NMT model. In baseline-TS, Step-1 is an English document generated by the Chinese-English NMT model, while Step-2 is an English headline generated by the English NHG model.

First, it is evident that the final results of the two baseline models are semantically different from the reference headlines, mainly because the two baseline models are pipeline methods and the errors of the first step are inevitably propagated to subsequent steps. In addition to the semantic differences, the headline generated by the baseline model has a greater length and does not meet the requirements of a concise headline. A comparison of the results of the two baseline models reveals that the quality of baseline-ST is slightly superior to that of baseline-TS due to shorter length and more similar semantics. This further verifies that if the data distribution of the training data of the two models in a pipeline method is significantly different, the problem of error propagation decreases the quality of the final headline.

When the reward function is $Reward\ 1$, “new government” appears in the system-generated headlines. Although “新 (new)” and “ 政府(government)” appear in the original input document, they do not appear together. Therefore, when using the generation probability of cross-lingual headlines as the reward function to train the model, although the model has the ability to capture key content “Cambodia,” because the reward function is ultimately defined at the word level it prevents the model from being able to capture global information.

When the reward function is $Reward\ 2$, although “new government” does not appear in the system-generated output, redundant content “khmer rouge” appears in the output of the pretraining model. The same problem does not occur in the reinforced training or joint training models, whose results are more consistent with the reference headline. Therefore, we conclude that the overall effect of the $Reward\ 2$ function is superior to that of the $Reward\ 1$ function.

Among the different training methods, the results of the pretraining method are always the poorest. For example, the redundant content “Khmer Rouge” appears in the results of $Reward\ 2$

Authorized licensed use limited to: Tsinghua University. Downloaded on December 16,2020 at 07:18:41 UTC from IEEE Xplore. Restrictions apply.

+ pretraining, and the redundant content "new government" also appears in the results of the $Reward\ 1$ + pretraining model (repeating "prince"). This problem can be somewhat corrected by reinforced training. The result of the $Reward\ 1$ + reinforced training model does not contain the repeated content "prince," and the result of the $Reward\ 2$ + reinforced training model does not contain the redundant content "Khmer Rouge" but contains the correct content "Prince Ranariddh." In this sample, the advantage of joint training is not particularly apparent; however, this may be consistent with the fact that the evaluation scores between joint training and reinforced training models in Table V do not differ significantly.

In summary, the above samples indicate that several model variants proposed in this study can improve the pipeline model from a certain aspect (headline length, semantics, or fluency), thus demonstrating that the construction of a direct model is of great important for CNHG.

V. Related Works

Our study builds on previous research in the field of NHG, cross-lingual summarization, and the use of reinforcement learning.

A. Neural Headline Generation

There has been remarkable progress in end-to-end NHG in recent years, and much research has been conducted to improve model performance from various perspectives. These research include augmenting the fixed-size vocabulary [34]–[36], tackling the online learning problem [37], addressing the unique characteristics of summarization [38]–[41], enhancing headline quality [42], [43], and integrating topic information [44].

B. Cross-Lingual Summarization

Cross-lingual summarization, which aims to help individuals capture the main theme of original news documents even if they are in another language, is of great significance in automatic summarization. In previous studies, researchers generally utilize statistical information or hand-crafted features to develop a cross-lingual summarization system. These studies include the two-step pipeline method of Wan *et al.* [45], the graph-based system of Wan [46], the machine-translation-based heuristic method of Yao *et al.* [47], and a multi-document summarization system that considers bilingual concepts and facts [48]. Benefitting from the end-to-end framework, there have been several recent attempts to model CNHG. Ayana *et al.* [9] propose using the teacher-student framework to construct a direct CNHG model, and Duan *et al.* [10] extende their research.

C. Use of Reinforcement Learning

Recent studies have demonstrated that reinforcement learning is an effective approach for improving the performance of a NHG system [21], [49]–[53]. For example, Paulus *et al.* [49] design a reinforcement learning method for a traditional NHG task, while Narayan *et al.* [21] and Dong *et al.* [50] both propose a neural-network-based extractive summarization model with reinforcement learning. The above studies all utilize ROUGE as the reward. In addition to ROUGE, [51] also introduce a ROUGE-related saliency score and entailment score to enhance the training procedure.

VI. Conclusion

In this study, we propose a reinforcement learning framework that is composed of two modules and attempts to utilize existing same-language headline generation training data and translation training data to address the cross-lingual headline generation problem. The fundamental principle is to use a translation module and cross-lingual headline generation module to form the entire reinforcement learning framework. The translation module translates the input document into the source language news document, which is then taken as input for the cross-lingual headline generation module. Then, the two modules are jointly trained through a reward related to the cross-lingual headline generation task. The experimental results on a Chinese-English cross-lingual headline generation task reveal that the proposed method significantly outperforms the baseline systems.

In future work, we plan to address several remaining problems. One obstacle in CNHG is that the training data from the Chinese NHG, English NHG, and Chinese-English NMT tasks remain imbalanced, as they are collected from different sources. We plan to automatically crawl appropriate news data, either in Chinese or English, from the web to address this problem. In addition, when superior training data are unavailable, it will be also interesting to explore how to utilize transfer learning methods to alleviate the existing problem.

References

[1] M. Banko, V. O. Mittal, and M. J. Witbrock, "Headline generation based on statistical translation," in *Proc. 38th Annu. Meeting Assoc. Comput. Linguist.*, 2000, pp. 318–325.

[2] B. Dorr, D. Zajic, and R. Schwartz, "Hedge trimmer: A parse-and-trim approach to headline generation," in *Proc. Human Lang. Technol. Conf. North Amer. Chapter Assoc. Comput. Linguist. Text Summarization Workshop*, 2003, pp. 1–8.

[3] M. Galley and K. McKeown, "Lexicalized Markov grammars for sentence compression," in *Proc. Human Lang. Technol.: The Conf. North Amer. Chapter Assoc. Comput. Linguist.*, 2007, pp. 180–187.

[4] T. Berg-Kirkpatrick, D. Gillick, and D. Klein, "Jointly learning to extract and compress," in *Proc. 49th Annu. Meeting Assoc. Comput. Linguist.: Human Lang. Technol.*, 2011, pp. 481–490.

[5] Z. Zhu, D. Bernhard, and I. Gurevych, "A monolingual tree-based translation model for sentence simplification," in *Proc. 23rd Int. Conf. Comput. Linguist.*, 2010, pp. 1353–1361.

[6] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," in *Proc. 32nd Int. Conf. Mach. Learn. Representations*, 2015.

[7] C. Napoles, M. Gormley, and B. Van Durme, "Annotated gigaword," in *Proc. Joint Workshop Autom. Knowl. Base Construction Web-scale Knowl. Extraction*, 2012, pp. 95–100.

[8] B. Hu, Q. Chen, and F. Zhu, "Lcsts: A large scale chinese short text summarization dataset," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2015, pp. 1967–1972.

[9] Ayana, S. Shen, Y. Chen, C. Yang, Z. Liu, and M. Sun, "Zero-shot cross-lingual neural headline generation," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 26, no. 12, pp. 2319–2327, Dec. 2018.

[10] X. Duan, M. Yin, M. Zhang, B. Chen, and W. Luo, "Zero-shot cross-lingual abstractive sentence summarization through teaching generation and attention," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguist.*, 2019, pp. 3162–3172.

Authorized licensed use limited to: Tsinghua University. Downloaded on December 16,2020 at 07:18:41 UTC from IEEE Xplore. Restrictions apply.

[11] J. D. Williams, K. Asadi, and G. Zweig, "Hybrid code networks: Practical and efficient end-to-end dialog control with supervised and reinforcement learning," in *Proc. 55th Annu. Meeting Assoc. Comput. Linguist.*, 2017, pp. 665–677.
[12] Z. Li, X. Jiang, L. Shang, and H. Li, "Paraphrase generation with deep reinforcement learning," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2018, pp. 3865–3878.
[13] J. Xu *et al.*, "Unpaired sentiment-to-sentiment translation: A cycled reinforcement learning approach," in *Proc. 56th Annu. Meeting Assoc. Comput. Linguist.*, 2018, pp. 979–988.
[14] D. Bahdanau *et al.*, "An actor-critic algorithm for sequence prediction," in *Proc. 5th Int. Conf. Learn. Representations*, 2017.
[15] L. Wu, F. Tian, T. Qin, J. Lai, and T.-Y. Liu, "A study of reinforcement learning for neural machine translation," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2018, pp. 3612–3621.
[16] R. J. Williams, "Simple statistical gradient-following algorithms for connectionist reinforcement learning," *Mach. Learn.*, vol. 8, no. 3, pp. 229–256, May 1992.
[17] D. He *et al.*, "Dual learning for machine translation," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 820–828.
[18] Y. Chen, Y. Liu, and V. Li, "Zero-resource neural machine translation with multi-agent communication game," in *Proc. 32nd AAAI Conf. Artif. Intell.*, 2018, pp. 5086–5093.
[19] S. Shen *et al.*, "Minimum risk training for neural machine translation," in *Proc. 54th Annu. Meeting Assoc. Comput. Linguist.*, 2016, pp. 1683–1692.
[20] M. Ranzato, S. Chopra, M. Auli, and W. Zaremba, "Sequence level training with recurrent neural networks," in *Proc. 4th Int. Conf. Learn. Representations*, 2016.
[21] S. Narayan, S. B. Cohen, and M. Lapata, "Ranking sentences for extractive summarization with reinforcement learning," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguist.: Human Lang. Technol.*, 2018, pp. 1747–1759.
[22] C. Sun *et al.*, "Extracting entities and relations with joint minimum risk training," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2018, pp. 2256–2265.
[23] Ayana, S. Shen, Y. Lin, C. Tu, and M. Sun, "Recent advances on neural headline generation," *J. Comput. Sci. Technol.*, vol. 32, no. 4, pp. 768–784, 2017.
[24] L. Yu, W. Zhang, J. Wang, and Y. Yu, "Seqgan: Sequence generative adversarial nets with policy gradient," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 2852–2858.
[25] L. Wu *et al.*, "Adversarial neural machine translation," 2017, *arXiv:1704.06933*.
[26] P. Li, L. Bing, and W. Lam, "Actor-critic based training framework for abstractive summarization," 2018, *arXiv:1803.11070*.
[27] M. D. Zeiler, "Adadelta: An adaptive learning rate method," *Comput. Sci.*, 2012, *arXiv:1212.5701*.
[28] T. T. D. Team, R. Al-Rfou, G. Alain, A. Almahairi, and Y. Zhang, "Theano: A Python framework for fast computation of mathematical expressions," 2016, *arXiv:1605.02688*.
[29] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "Bleu: a method for automatic evaluation of machine translation," in *Proc. 40th Annu. Meeting Assoc. Comput. Linguist.*, 2002, pp. 311–318.
[30] C.-Y. Lin, "Rouge: A package for automatic evaluation of summaries," in *Proc. 42nd Annu. Meeting Assoc. Comput. Linguist. Text Summarization Branches Out Workshop*, 2004, vol. 8, pp. 74–81.
[31] A. Vaswani *et al.*, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.
[32] Y. Chen, Y. Liu, Y. Cheng, and O. V. Li, "A teacher-student framework for zero-resource neural machine translation," in *Proc. 55th Annu. Meeting Assoc. Comput. Linguist.*, 2017, pp. 1925–1935.
[33] Y. Kim and A. M. Rush, "Sequence-level knowledge distillation," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2016, pp. 1317–1327.
[34] Z. Cao, C. Luo, W. Li, and S. Li, "Joint copying and restricted generation for paraphrase," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 3152–3158.
[35] J. Gu, Z. Lu, H. Li, and V. O. Li, "Incorporating copying mechanism in sequence-to-sequence learning," in *Proc. 54th Annu. Meeting Assoc. Comput. Linguist.*, 2016, pp. 1631–1640.
[36] C. Gulcehre, S. Ahn, R. Nallapati, B. Zhou, and Y. Bengio, "Pointing the unknown words," in *Proc. 54th Annu. Meeting Assoc. Comput. Linguist.*, 2016, pp. 140–149.
[37] L. Yu, J. Buys, and P. Blunsom, "Online segment to segment neural transduction," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2016, pp. 1307–1316.
[38] Y. Kikuchi, G. Neubig, R. Sasano, H. Takamura, and M. Okumura, "Controlling output length in neural encoder-decoders," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2016, pp. 1328–1338.
[39] P. Li, W. Lam, L. Bing, and Z. Wang, "Deep recurrent generative decoder for abstractive text summarization," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2017, pp. 2091–2100.
[40] Y. Miao and P. Blunsom, "Language as a latent variable: Discrete generative models for sentence compression," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2016, pp. 319–328.
[41] Q. Zhou, N. Yang, F. Wei, and M. Zhou, "Selective encoding for abstractive sentence summarization," in *Proc. 55th Annu. Meeting Assoc. Comput. Linguist.*, 2017, pp. 1095–1104.
[42] Z. Cao, W. Li, S. Li, and F. Wei, "Retrieve, rerank and rewrite: Soft template based neural summarization," in *Proc. 56th Annu. Meeting Assoc. Comput. Linguist.*, 2018, pp. 152–161.
[43] Z. Cao, F. Wei, W. Li, and S. Li, "Faithful to the original: Fact aware neural abstractive summarization," in *Proc. 32nd AAAI Conf. Artif. Intell.*, 2018, pp. 4784–4791.
[44] L. Wang, J. Yao, Y. Tao, L. Zhong, W. Liu, and Q. Du, "A reinforced topic-aware convolutional sequence-to-sequence model for abstractive text summarization," in *Proc. Int. Joint Conf. Artif. Intell.*, 2018, pp. 4453–4460.
[45] X. Wan, H. Li, and J. Xiao, "Cross-language document summarization based on machine translation quality prediction," in *Proc. 48th Annu. Meeting Assoc. Comput. Linguist.*, 2010, pp. 917–926.
[46] X. Wan, "Using bilingual information for cross-language document summarization," in *Proc. 49th Annu. Meeting Assoc. Comput. Linguist.*, 2011, pp. 1546–1555.
[47] J.-g. Yao, X. Wan, and J. Xiao, "Phrase-based compressive cross-language summarization," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2015, pp. 118–127.
[48] J. Zhang, Y. Zhou, and C. Zong, "Abstractive cross-language summarization via translation model enhanced predicate argument structure fusing," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 10, pp. 1842–1853, Oct. 2016.
[49] R. Paulus, C. Xiong, and R. Socher, "A deep reinforced model for abstractive summarization," in *Proc. 5th Int. Conf. Learn. Representations*, 2018, *arXiv:1705.04304*.
[50] Y. Dong, Y. Shen, E. Crawford, H. van Hoof, and J. C. K. Cheung, "Banditsum: Extractive summarization as a contextual bandit," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2018, pp. 3739–3748.
[51] R. Pasunuru and M. Bansal, "Multi-reward reinforced summarization with saliency and entailment," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguist.: Human Lang. Technol.*, 2018, pp. 646–653.
[52] Y.-C. Chen and M. Bansal, "Fast abstractive summarization with reinforce-selected sentence rewriting," in *Proc. 56th Annu. Meeting Assoc. Comput. Linguist.*, 2018, pp. 675–686.
[53] Y. Wu and B. Hu, "Learning to extract coherent summary via deep reinforcement learning," in *Proc. 32nd AAAI Conf. Artif. Intell.*, 2018, pp. 5602–5609.

Ayana received the B.E. and M.E. degrees from the College of Computer Science and Technology, Inner Mongolia University, Hohhot, China, in 2006 and 2009, respectively. She received the Ph.D. degree from Tsinghua University in 2019. She is a Lecturer in Department of Computer Information Management, Inner Mongolia University of Finance and Economics, Hohhot, China. Her research interests include natural language processing, especially document summarization.

Authorized licensed use limited to: Tsinghua University. Downloaded on December 16,2020 at 07:18:41 UTC from IEEE Xplore. Restrictions apply.

Yun Chen received the B.E. degree from Tsinghua University in 2013 and the Ph.D. degree from the University of Hong Kong in 2018. She is an Assistant Professor in School of Information Management & Engineering, Shanghai University of Finance and Economics. Her research interests include machine learning approaches that are both linguistically motivated and tailored to natural language processing, especially neural machine translation.

Cheng Yang received the B.E. and Ph.D. degrees from Tsinghua University in 2014 and 2019, respectively. He is an Assistant Professor in Beijing University of Posts and Telecommunications. His research interests include natural language processing and network representation learning. He has published several top-level papers in international journals and conferences including ACM TOIS, EMNLP, IJCAI and AAAI.

Zhiyuan Liu received the B.Eng. degree in 2006 and the Ph.D. degree in 2011 from the Department of Computer Science and Technology, Tsinghua University. He is an Associate Professor of the Department of Computer Science and Technology, Tsinghua University. His research interests are natural language processing and social computation. He has published over 40 papers in international journals and conferences including ACM Transactions, IJCAI, AAAI, ACL and EMNLP.

Maosong Sun received the B.Eng. degree in 1986 and M.Eng. degree in 1988 from Department of Computer Science and Technology, Tsinghua University, and the Ph.D. degree in 2004 from Department of Chinese, Translation, and Linguistics, City University of Hong Kong. He is a professor of the Department of Computer Science and Technology, Tsinghua University. His research interests include natural language processing, Chinese computing, Web intelligence, and computational social sciences. He has published over 150 papers in academic journals and international conferences in the above fields. He serves as a Vice President of the Chinese Information Processing Society, the Council Member of China Computer Federation, the Director of Massive Online Education Research Center of Tsinghua University, and the Editor-in-Chief of the Journal of Chinese Information Processing.

Authorized licensed use limited to: Tsinghua University. Downloaded on December 16,2020 at 07:18:41 UTC from IEEE Xplore. Restrictions apply.