System Description on Third Automatic Simultaneous Translation Workshop

Zhang Yiqiao
College of science, Huazhong Agricultural University, Wuhan 430070, China
qiaoyizhang@webmail.hzau.edu.cn

Abstract

This paper shows my submission to the Third Automatic Simultaneous Translation Workshop at NAACL2022. The submission includes Chinese audio to English text task, Chinese text to English text tast, and English text to Spanish text task. For the two text-to-text tasks, I use the STACL model of PaddleNLP. As for the audio-to-text task, I first use DeepSpeech2 to translate the audio into text, then apply the STACL model to handle the text-to-text task. The submission results show that the used method can get low delay with a few training samples.

1 Introduction

The submitted system consists of two parts. One is audio to text system, which can translate Chinese audio into English text. The second part is the text-to-text model, which can translate source text into the target language.

In the text-to-text translation task, the used system is STACL model (Ma et al., 2019). All training data are processed by Byte Pair Encoding (Sennrich et al., 2016). In addition, the strategies used by the model in training and inference are the same. For example, if the wait-k strategy in inference is 1, the wait-k in training is also 1.

In the audio to text translation task, the DeepSpeech2 model (Amodei et al., 2015) is used as the preprocessing of the STACL model. The DeepSpeech2 model can translate audio (Chinese) segments into text (Chinese) segments and then input the segments into the STACL model to generate the target-language text.

The submitted results show that the used STACL model has a low delay for text translation tasks. But the system can only generate the results with a high delay in the audio translation task.

The rest of the paper is organized as follows. Section 2 describes the training data used in the submitted system. Section 3 describes the model, training strategy, and results. The conclusions are given in Section 4.

2 Datasets

In this section, I describe the Datasets.

2.1 Zh-En Text Translation Dataset

The dataset used for the Chinese-to-English(Zh-En) translation task is extracted from AST, which is provided by the NAACL workshop. This data set contains 214 JSON files, and each JSON file contains parallel Chinese vs. English corpus. The data, which is extracted from these JSON files, contains 37,901 Chinese vs. English samples. After byte pair encoding, the samples are used to train the Zh-En translation model.

The BPE vocabulary of the Zh-En translation task can be found in the Github project of PaddleNLP (Contributors, 2021).

2.2 En-Es Text Translation Dataset

The dataset used for the English-to-Spanish(En-Es) text translation task was obtained from the United Nations Parallel Corpus(Ziemski et al., 2016). The En-Es dataset contains 21,911,121 samples. After byte pair encoding, the dataset is used to train the En-Es text translation model.

For obtaining the BPE vocabulary, I segment the source dataset into subword units by Subword Neural Machine Translation (Sennrich et al., 2015). The code for segmentation can be found in (Sennrich, 2021).

2.3 Audio-to-Test Dataset

The training data of the Chinese speech recognition model is AISHELL (Bu et al., 2017), which is an open-source Mandarin speech corpus. In the submitted system, I only use the pre-trained model of the DeepSpeech2 model on AISHELL.

Proceedings of the Third Workshop on Automatic Simultaneous Translation, pages 22 - 24

Parameter	Value
wait-k	1 or 3
max epoch	30
batch size	512
learning rate	2.0
max length	256
n layer	6

Table 1: Training parameters in Zh-En translation model

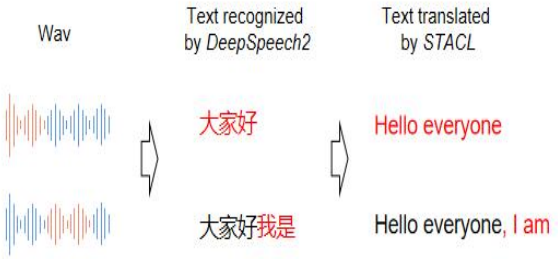

Figure 1: Frame for audio translation system

3 Models and Results

This section shows the models used in the submitted system and discusses the results.

3.1 Text Translation System

3.1.1 STACL model

In the text translation task, the model is STACL (Ma et al., 2019), which is a translation architecture for all simultaneous interpreting scenarios. For train the model, the wait-k strategy is adopted. The model will wait for k words of the source text and then start to translate. For example, when k is 2, the model only starts translating the first word of the target language after obtaining the second word of the source text.

In the inference process, the model decodes one word at a time. When the sentences to be translated are all read, the untranslated sentences will be completed at once.

3.1.2 Results in Zh-En task

In the Zh-En translation task, I trained the model with wait-k = 1 and wait-k = 3. The details of training parameters are shown in table 1.

When the wait-k is 1, the AL of the submitted result is -1.28, and the BLEU is 14.86. When the wait-k is 3, the AL is -0.52, and the BLUE is 14.84. The two results have almost the same accuracy, demonstrating that the used dataset may not be sufficient for the translation task.

3.1.3 Results in En-Es task

In the En-Es translation task, the max epoch is set as 1, and other parameters are the same as table 1.

The AL of the submitted result is -1.61, and the BLEU is 11.82.

3.2 Audio Translation System

3.2.1 DeepSpeech2 model

Deepspeech2 is an end-to-end automatic speech recognition system based on the PaddlePaddle deep learning framework (Amodei et al., 2015). In order to translate the speech data into the corresponding target-language text, I first segment the audio, use deepspeech2 to covert the voice segment into text, and then translate the recognized text into the target language through the STACL model. Figure 1 shows the workflow of speech recognition translation.

3.2.2 Results

Since each segment contains multiple Chinese characters, decoding only one character at a time will lead to excessive delay (CW value). To overcome this issue, I decoded two characters at once. The CW of submitted results is 19.21, and the BLEU is 7.3.

4 Conclusion

This paper describes my submitted system at the Third Automatic Simultaneous Translation Workshop. The system submitted has a low delay. I will conduct a further study about the speech recognition strategy in the future.

Acknowledgements

This document has been adapted by Steven Bethard, Ryan Cotterell and Rui Yan from the instructions for earlier ACL and NAACL proceedings, including those for ACL 2019 by Douwe Kiela and Ivan Vulić, NAACL 2019 by Stephanie Lukin and Alla Roskovskaya, ACL 2018 by Shay Cohen, Kevin Gimpel, and Wei Lu, NAACL 2018 by Margaret Mitchell and Stephanie Lukin, BibTEX suggestions for (NA)ACL 2017/2018 from Jason Eisner, ACL 2017 by Dan Gildea and Min-Yen Kan, NAACL 2017 by Margaret Mitchell, ACL 2012 by Maggie Li and Michael White, ACL 2010 by Jing-Shin Chang and Philipp Koehn, ACL 2008 by Johanna D. Moore, Simone Teufel, James Allan, and Sadaoki Furui, ACL 2005 by Hwee Tou Ng and

Kemal Oflazer, ACL 2002 by Eugene Charniak and Dekang Lin, and earlier ACL and EACL formats written by several people, including John Chen, Henry S. Thompson and Donald Walker. Additional elements were taken from the formatting instructions of the *International Joint Conference on Artificial Intelligence* and the *Conference on Computer Vision and Pattern Recognition*.

References

Dario Amodei, Rishita Anubhai, Eric Battenberg, Carl Case, Jared Casper, Bryan Catanzaro, Jingdong Chen, Mike Chrzanowski, Adam Coates, Greg Diamos, Erich Elsen, Jesse H. Engel, Linxi Fan, Christopher Fougner, Tony Han, Awni Y. Hannun, Billy Jun, Patrick LeGresley, Libby Lin, Sharan Narang, Andrew Y. Ng, Sherjil Ozair, Ryan Prenger, Jonathan Raiman, Sanjeev Satheesh, David Seetapun, Shubho Sengupta, Yi Wang, Zhiqian Wang, Chong Wang, Bo Xiao, Dani Yogatama, Jun Zhan, and Zhenyao Zhu. 2015. Deep speech 2: End-to-end speech recognition in english and mandarin. *CoRR*, abs/1512.02595.

Hui Bu, Jiayu Du, Xingyu Na, Bengu Wu, and Hao Zheng. 2017. AISHELL-1: an open-source mandarin speech corpus and A speech recognition baseline. *CoRR*, abs/1709.05522.

PaddleNLP Contributors. 2021. Paddlenlp: An easy-to-use and high performance nlp library. https://github.com/PaddlePaddle/PaddleNLP.

Mingbo Ma, Liang Huang, Hao Xiong, Renjie Zheng, Kaibo Liu, Baigong Zheng, Chuanqiang Zhang, Zhongjun He, Hairong Liu, Xing Li, Hua Wu, and Haifeng Wang. 2019. STACL: Simultaneous translation with implicit anticipation and controllable latency using prefix-to-prefix framework. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3025–3036, Florence, Italy. Association for Computational Linguistics.

Rico Sennrich. 2021. Subword neural machine translation. https://github.com/rsennrich/subword-nmt.

Rico Sennrich, Barry Haddow, and Alexandra Birch. 2015. Neural machine translation of rare words with subword units. *CoRR*, abs/1508.07909.

Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Neural machine translation of rare words with subword units. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1715–1725, Berlin, Germany. Association for Computational Linguistics.

Michał Ziemski, Marcin Junczys-Dowmunt, and Bruno Pouliquen. 2016. The united nations parallel corpus v1.0.