Automatic Translating between Ancient Chinese and Contemporary Chinese with Limited Aligned Corpora

Zhiyuan Zhang[1], Wei Li[1], and Qi Su[2]

[1] MOE Key Lab of Computational Linguistics, School of EECS, Peking University
[2] School of Foreign Languages, Peking University
{zzy1210, liweitj47, sukia}@pku.edu.cn

Abstract. The Chinese language has evolved a lot during the long-term development. Therefore, native speakers now have trouble in reading sentences written in ancient Chinese. In this paper, we propose to build an end-to-end neural model to automatically translate between ancient and contemporary Chinese. However, the existing ancient-contemporary Chinese parallel corpora are not aligned at the sentence level and sentence-aligned corpora are limited, which makes it difficult to train the model. To build the sentence level parallel training data for the model, we propose an unsupervised algorithm that constructs sentence-aligned ancient-contemporary pairs by using the fact that the aligned sentence pair shares many of the tokens. Based on the aligned corpus, we propose an end-to-end neural model with copying mechanism and local attention to translate between ancient and contemporary Chinese. Experiments show that the proposed unsupervised algorithm achieves 99.4% F1 score for sentence alignment, and the translation model achieves 26.95 BLEU from ancient to contemporary, and 36.34 BLEU from contemporary to ancient.

Keywords: Sentence Alignment · Neural Machine Translation

1 Introduction

The ancient Chinese was used for thousands of years. There is a huge amount of books and articles written in ancient Chinese. However, both the form and grammar of the ancient Chinese have been changed. Chinese historians and littérateurs have made great efforts in translating such literature into contemporary Chinese, a big part of which is publicly available on the Internet. However, there is still a big gap between these literatures and parallel corpora, because most of the corpora are coarsely passage-aligned, the orders of sentences are different. To train an automatic translating model, we need to build a sentence-aligned corpus first.

Translation alignment is an important pre-step for machine translation. Most of the previous work focuses on how to apply supervised algorithms on this task using features extracted from texts. The Gale Church algorithm [5, 7] uses statistical or dictionary information to build alignment corpus. Strands [18, 19] extracts parallel corpus from the Internet. A logarithmic linear model [26] is also used for Chinese-Japanese clause alignment. Besides, features such as sentence lengths, matching patterns, Chinese character

co-occurrence in Japanese and Chinese are also taken into consideration. The observation that Chinese character co-occurrence also exists in ancient-contemporary Chinese is also used to ancient-contemporary Chinese translation alignment [11, 12].

The method above works well, however, these supervised algorithms require a large parallel corpus to train, which is not available in our circumstance. These previous algorithms did not make good use of the characteristics of ancient-contemporary Chinese pair. To overcome these shortcomings, we design an unsupervised algorithm for sentence alignment based on the observation that differently from a bilingual corpus, ancient-contemporary sentence pairs share many common characters in order. We evaluate our alignment algorithm on an aligned parallel corpus with small size. The experimental results show that our simple algorithm works very well (F1 score 99.4), which is even better than the supervised algorithms.

Deep learning has achieved great success in tasks like machine translation. The sequence to sequence (seq-to-seq) model [22] is proposed to generate good translation results on machine translation. The attention mechanism [1] is proposed to allow the decoder to extract phrase alignment information from the hidden states of the encoder. Most of the existing NMT systems are based on the seq-to-seq model [9, 3, 23] and the attention mechanism. Some of them have variant architectures to capture more information from the inputs [21, 25], and some improve the attention mechanism [13, 16, 17, 8, 4, 2], which also enhanced the performance of the NMT model. Inspired by the work of pointer generator [20], we use a copying mechanism named pointer-generator model to deal with the task of ancient-contemporary translation task because ancient and contemporary Chinese share common characters and some same proper nouns.

Experimental results show that a copying mechanism can improve the performance of seq-to-seq model remarkably on this task. We show some experimental results in the experiment section. Other mechanisms are also implied to improve the performance of machine translation [10, 15].

Our contributions lie in the following two aspects:

- We propose a simple yet effective unsupervised algorithm to build the sentence-aligned parallel corpora, which make it possible to train an automatic translating model between ancient Chinese and contemporary Chinese even with limited sentence-aligned corpora.
- We propose to apply the sequence to sequence model with copying mechanism and local attention to deal with the translating task. Experimental results show that our method can achieve the BLEU score of 26.41 (ancient to contemporary) and 35.66 (contemporary to ancient).

2 Proposed Method

2.1 Unsupervised Algorithm for Sentence Alignment

We review the definition of sentence alignment first.

Given a pair of aligned passages, the source language sentences S and target language sentences T, which are defined as

$$S := \{s_1, s_2, \cdots, s_n\}, \quad T := \{t_1, t_2, \cdots, t_m\} \tag{1}$$

Algorithm 1 Dynamic Programming Algorithm to Find the Alignment Result.

Preprocess **lcs** scores and store them in memory. Initialize $f[i,j] \leftarrow 0$.
for $i \in [1,n]$
  for $j \in [1,m]$
    for $i', j' \in \{(i',j') : \{i-i', j-j'\} = \{1,M\}, 0 \leq M \leq 5, i' > 0, j' > 0\}$
      Get $now\_lcs \leftarrow$ **lcs**$(s[i'+1,i], t[j'+1,j])$ from memory.
      if $f[i',j'] + now\_lcs > f[i,j]$
        $f[i,j] = f[i',j'] + now\_lcs$
        Update the corresponding alignment result.
      endif
    endfor
  endfor
endfor

The objective of sentence alignment is to extract a set of matching sentence pairs out of the two passages. Each matching pair consists of several sentence pairs like (s_i, t_j), which implies that s_i and t_j form a parallel pair.

For instance, if $S = \{s_1, s_2, s_3, s_4, s_5\}, T = \{t_1, t_2, t_3, t_4, t_5\}$ and the alignment result is $\{(s_1,t_1), (s_3,t_2), (s_3,t_3), (s_4,t_4), (s_5,t_4)\}$. It means source sentence s_1 is translated to target sentence t_1, s_3 is translated to t_2 and t_3, s_4 and s_5 are translated to t_4 while s_2 is not translated to any sentence in target sentences. Here we may assume that source sentences are translated to target sentences in order. For instance (s_1, t_2) and (s_2, t_1) can not exist at the same time in the alignment result.

Translating ancient Chinese into contemporary Chinese has the characteristic that every word of ancient Chinese tends to be translated into contemporary Chinese in order, which usually includes the same original character. Therefore, correct aligned pairs usually have the maximum sum of lengths of the longest common subsequence (LCS) for each matching pair.

Let **lcs**$(s[i_1,i_2], t[j_1,j_2])$ be the length of the **longest common subsequence** of a matching pair of aligned sentences consisting of source language sentences $s[i_1,i_2]$ and target language sentences $t[j_1,j_2]$, which are defined as

$$s[i_1,i_2] = \begin{cases} s_{i_1}s_{i_1+1}\cdots s_{i_2} (i_1 \leq i_2) \\ [empty\_str]\ (i_1 > i_2) \end{cases}, \quad t[j_1,j_2] = \begin{cases} t_{j_1}t_{j_1+1}\cdots t_{j_2} (j_1 \leq j_2) \\ [empty\_str]\ (j_1 > j_2) \end{cases} \tag{2}$$

where $[empty\_str]$ denotes an empty string. The longest common subsequence of an empty string and any string is 0.

We use the dynamic programming algorithm to find the maximum score and its corresponding alignment result. Let $f(i,j)$ be the maximum score that can be achieved with partly aligned sentence pairs until s_i, t_j. To reduce the calculation cost, we only consider cases where one sentence is matched with no more than 5 sentences:

$$f(i,j) = \max_{\{i-i', j-j'\} = \{1,M\}, 0 \leq M \leq 5} \{f(i',j') + \textbf{lcs}(s[i'+1,i], t[j'+1,j])\} \tag{3}$$

The condition $\{i-i', j-j'\} = \{1,M\}, 0 \leq M \leq 5$ ensures that one sentence is matched with no more than 5 sentences and $M = 0$ holds if and only if one sentence

is not matched with any of the sentences. We can preprocess all $\mathbf{lcs}(s[i'+1,i],t[j'+1,j])(\{i-i',j-j'\}=\{1,M\},0\leq M\leq 5)$ scores and store them for using in algorithm, which has a time complexity of $O(mn)$. The pseudo code is shown in Algorithm 1. Then for every $f(i,j)$, we only need to enumerate i',j' for $O(1)$ times and the time complexity of proposal dynamic programming algorithm is $O(mn)$.

When the size of corpus grows, this algorithm will be time-consuming, which means our algorithm is more suitable for passage-aligned corpora instead of a huge text. In reality, we find that when translating a text written in ancient Chinese into contemporary Chinese. Translators tend not to change the structure of the text, namely, a book written in ancient Chinese and its contemporary Chinese version are a passage-aligned corpus naturally. Therefore, we can get abundant passage-aligned corpora form Internet.

2.2 Neural Machine Translation Model

Sequence-to-sequence model was first proposed to solve machine translation problem. The model consists of two parts, an encoder and a decoder. The encoder is bound to take in the source sequence and compress the sequence into hidden states. The decoder is used to produce a sequence of target tokens based on the information embodied in the hidden states given by the encoder. Both encoder and decoder are implemented with Recurrent Neural Networks (RNN).

To deal with the ancient-contemporary translating task, we use the encoder to convert the variable-length character sequence into h_t, the hidden representations of position t, with an Bidirectional RNN:

$$\overrightarrow{h_t}=f(x_t,\overrightarrow{h_{t-1}}),\quad \overleftarrow{h_t}=f(x_t,\overleftarrow{h_{t+1}}),\quad h_t=[\overrightarrow{h_t};\overleftarrow{h_t}] \tag{4}$$

Where f is a function of RNN family, x_t is the input at time step t. The decoder is another RNN, which generates a variable-length sequence token by token, through a conditional language model,

$$s_t=f(c_t,s_{t-1},Ey_{t-1}),\quad c_t=g(\boldsymbol{h},s_{t-1}) \tag{5}$$

Where E is the embedding matrix of target tokens, y_{t-1} is the last predicted token. We also implement a beam search mechanism, a heuristic search algorithm that expands the most promising node in a limited set, for generating a better target sentence.

In the decoder, suppose l is the length of the source sentence, then the context vector c_t is calculated based on the hidden states s_t of the decoder at time step t and all the hidden states $\boldsymbol{h}=\{h_i\}_{i=1}^{l}$ in the encoder, which is also known as the attention mechanism,

$$\beta_{t,i}=v_a^{\mathrm{T}}\tanh(W_a s_t+U_a h_i+b_a),\quad a_{t,i}=\frac{\exp(\beta_{t,i})}{\sum_{j=1}^{l}\exp(\beta_{t,j})} \tag{6}$$

We adopt the global attention mechanism in the baseline seq-to-seq model,

$$c_t=\sum_{i=1}^{l}a_{t,i}h_i \tag{7}$$

where $a_{t,i}$ is the attention probability of the word in position i of the source sentence in the decoder time step t and h_i is the hidden state of position i of the source sentence.

Because in most cases ancient and contemporary Chinese have similar word order, instead of the normal global attention in the baseline seq-to-seq model, we apply *local attention* [14, 24] in our proposal. When calculating the context vector c_t, we calculate a pivot position in the hidden states $\boldsymbol{h}$ of the encoder, and calculate the attention probability in the window around the pivot instead of the whole sentence,

$$c_t = \sum_{i=1}^{l} p_{t,i} a_{t,i} h_i \tag{8}$$

where $p_{t,i}$ is the mask score based on the pivot position in the hidden states $\boldsymbol{h}$ of the encoder.

Machine translation model treats ancient and contemporary Chinese as two languages, however, in this task, contemporary and ancient Chines share many common characters. Therefore, we treat ancient and contemporary Chinese as one language and share the character embedding between the source language and target language.

2.3 Copying Mechanism

As is stated above, ancient and contemporary Chinese share many common characters and most of the name entities use the same representation. Copying mechanism [6] is very suitable in this situation, where the source and target sequence share some of the words. We apply pointer-generator framework in our model, which follows the same intuition as the copying mechanism. $p_t(w)$, the output probability of token w is calculated as,

$$p_t(w) = p_t^G p_t^V(w) + (1 - p_t^G) \sum_{\text{src}_i = w} a_{t,i}, \quad p_t^G = \sigma(W_g s_t + U_g c_t + b_g) \tag{9}$$

where p_t^G is dynamically calculated based on the hidden state s_t and the context vector c_t, $P_t^V(w)$ is the probability for token w in traditional seq-to-seq model. $a_{t,i}$ is the attention probability at the time step t in decoder and position i in source sentence which satisfies $\text{src}_i = w$, namely the token in position i in source sentence is w. σ is the sigmoid function.

The encoder and decoder networks are trained jointly to maximize the conditional probability of the target sequence. We use cross-entropy as the loss function. We use characters instead of words because characters have independent meaning in ancient Chinese and the number of characters is much lower than the number of words, which makes the data less sparse and greatly reduces the number of OOV. We also can implement the pointer-generator mechanism in summarization on our AT-seq-to-seq model to copy some proper nouns directly from the source language.

Table 1. Vocabulary statistics. We include all the characters in the training set in the vocabulary.

Language	Vocabulary	OOV Rate
Ancient	5,870	1.37%
Contemporary	4,993	1.03%

3 Experiments

3.1 Datasets

Sentence Alignment We crawl passages and their corresponding contemporary Chinese version from the Internet. To guarantee the quality of the contemporary Chinese translation, we choose the corpus from two genres, classical articles and Chinese historical documents. After proofreading a sample of these passages, we think the quality of the **passage-aligned** corpus is satisfactory. To evaluate the algorithm, we crawl a relatively small **sentence-aligned** corpus consisting of 90 aligned passages with 4,544 aligned sentence pairs. We proofread them and correct some mistakes to guarantee the correctness of this corpus.

Translating We conduct experiments on the data set built by our proposed unsupervised algorithm. The data set consists of 57,391 ancient-contemporary Chinese sentence pairs in total. We split sentence pairs randomly into train/dev/test dataset with sizes of 53,140/2,125/2,126 respectively. The vocabulary statistics information is in Table 1.

3.2 Experimental Settings

Sentence Alignment We implement a log-linear model on contemporary-ancient Chinese sentence alignment as a baseline. Following the previous work, we implement this model with a combination of three features, sentence lengths, matching patterns and Chinese character co-occurrence [26, 11, 12]. We split the data into training set (2,999) and test set (1,545) to train the log-linear model. Our unsupervised method does not need training data. Both these two methods are evaluated on the test set.

Translating We conduct experiments on both translating directions and use BLEU score to evaluate our model. We implement a one-layer Bidirectional LSTM with a 256-dim embedding size, 256-dim hidden size as encoder and a one-layer LSTM with attention mechanism as decoder. We also implement local attention or pointer generator mechanism on the decoder in our proposed model. We adopt Adam optimizer to optimize our loss functions with a learning rate of 0.0001. The training batch size of our model is 64 and the generating batch size is 32. The source language and target language share the vocabulary.

3.3 Experimental Results

Sentence Alignment Our unsupervised algorithm gets an F1-score of 99.4%, which is better than the supervised baseline with all features, 99.2% (shown in Table 2). Our

Table 2. Evaluation of logarithmic linear models and our method.

Features	Precision	Recall	F1-score
Length	0.821	0.855	0.837
Pattern	0.331	0.320	0.325
Length and Pattern	0.924	0.912	0.918
Co-occurrence	0.982	0.984	0.983
Length and Co-occurrence	0.987	0.989	0.988
Pattern and Co-occurrence	0.984	0.980	0.982
All features	0.992	0.991	0.992
our method	**0.994**	**0.994**	**0.994**

Table 3. Evaluation result (BLEU) of translating between **An** (ancient) and **Con** (contemporary) Chinese in test dataset

Method	An-Con	Con-An
Seq-to-Seq	23.10	31.20
+ copying	26.41	35.66
+ Local Attention	**26.95**	**36.34**

proposal algorithm does not use the features of sentence lengths and matching pattern. If we only adopt the feature of Co-occurrence, the performance of our proposal is 1.1% higher on F1-score than that of the supervised baseline (98.3%). To conclude, our proposal algorithm uses fewer features and use no training data but perform better than the supervised baseline.

Translating Experimental results show our model works well on this task (Table 3). Compared with the basic seq-to-seq model, copying mechanism gives a large improvement, because the source sentences and target sentence share some common representations, we will also give an example in the case study section. Local attention gives a small improvement over the traditional global attention, this can be attributed to shrinking the attention range, because most of the time, the mapping between ancient and contemporary Chinese is clear. A more sophisticated attention mechanism which makes full use of the characteristics of ancient and contemporary Chinese may further improve the performance.

4 Discussion

4.1 Sentence Alignment

We find that the small fraction of data (0.6%) that our method makes mistakes are mainly because of the change of punctuation. For example, in ancient Chinese, there is a comma "," after "异哉，" (How strange!), while in contemporary Chinese, "怪啊！" (How strange!), an exclamation mark "!" is used, which makes the sentence to be an

Table 4. Evaluation result (BLEU) of translating between **An** (ancient) and **Con** (contemporary) Chinese with different number of training samples.

Language	5,000	10,000	20,000	53,140
An-Con	3.00	9.69	16.31	26.95
Con-An	2.40	10.14	18.47	36.34

Table 5. Example of translating from **An** (ancient) to **Con** (contemporary) Chinese.

Source	六月辛卯，中山王焉薨。
(Translation in English)	(On Xinmao Day of the sixth lunar month, Yan, King of Zhongshan, passed away.)
Target	六月十二日，中山王刘焉逝世。
(Translation in English)	(On twelfth of the sixth lunar month, Liu Yan, King of Zhongshan, passed away.)
Seq2Seq	六月十六日，中山王刘裕去世。
(Translation in English)	(On sixteenth of the sixth lunar month, Liu Yu, King of Zhongshan, died.)
Our model	六月二十二日，中山王刘焉逝世。
(Translation in English)	(On twenty-second of the sixth lunar month, Liu Yan, King of Zhongshan, passed away.)

independent sentence. Since the sentence is short and there is no common character, our method fails to align the sentences correctly. However, such problems also exist in supervised models.

4.2 Necessity of Building a Large Sentence-aligned Corpus

From Table 4, we can see that the results are very sensitive to the scale of the training data size. Therefore, our unsupervised method of building a large sentence-aligned corpus is necessary. If we do not build a large sentence-aligned corpus by our sentence alignment algorithm, we will only have limited sentence-aligned corpora and the performance of translating will be worse.

4.3 Translating Result

Under most circumstances, our models can translate sentences between ancient Chinese and contemporary Chinese properly. For instance in Table 5, our models can translate "薨 (pass away)" into "去世 (pass away)" or "逝世 (pass away)", which are the correct forms of expression in contemporary Chinese. And our models can even complete some omitted characters. For instance, the family name "刘 (Liu)" in "中山王刘焉 (Liu Yan, King of Zhongshan)" was omitted in ancient Chinese because "中山王 (King of Zhongshan)" was a hereditary peerage offered to "刘 (Liu)" family. And our model completes the family name "刘 (Liu)" when translating. For proper nouns, the seq-to-2seq baseline model can fail sometimes while the copying model can correctly copy

them from the source language. For instance, the seq-to-seq baseline model translates "焉 (Yan)" into "刘裕 (Liu Yu, a famous figure in the history)" because "焉 (Yan)" is relatively low-frequent words in ancient Chinese. However, the copying model learns to copy these low-frequency proper nouns from source sentences directly.

Translating dates between ancient and contemporary Chinese calendar requires background knowledge of the ancient Chinese lunar calendar, and involves non-trivial calculation that even native speakers cannot translate correctly without training. In the example, "辛卯 (The Xinmao Day)" is the date presented in the ancient form, our model fails to translate it. Our model fails to transform between the Gregorian calendar and the ancient Chinese lunar calendar and choose to generate a random date, which is expected because of the difficulty of such problems.

5 Conclusion

In this paper, we propose an unsupervised algorithm to construct sentence-aligned sentence pairs out of a passage-aligned corpus to build a large sentence-aligned corpus. We propose to apply the sequence to sequence model with attention and copying mechanism to automatically translate between two styles of Chinese sentences. The experimental results show that our method can yield good translating results.

References

1. Bahdanau, D., Cho, K., Bengio, Y.: Neural machine translation by jointly learning to align and translate. In: Bengio, Y., LeCun, Y. (eds.) 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings (2015)
2. Calixto, I., Liu, Q., Campbell, N.: Doubly-attentive decoder for multi-modal neural machine translation. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers. pp. 1913–1924 (2017)
3. Cho, K., van Merrienboer, B., Gülçehre, Ç., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y.: Learning phrase representations using RNN encoder-decoder for statistical machine translation. In: EMNLP 2014. pp. 1724–1734 (2014)
4. Feng, S., Liu, S., Yang, N., Li, M., Zhou, M., Zhu, K.Q.: Improving attention modeling with implicit distortion and fertility for machine translation. In: COLING 2016. pp. 3082–3092 (2016)
5. Gale, W.A., Church, K.W.: A program for aligning sentences in bilingual corpora. Computational Linguistics **19**, 75–102 (1993)
6. Gu, J., Lu, Z., Li, H., Li, V.O.K.: Incorporating copying mechanism in sequence-to-sequence learning. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 1: Long Papers (2016)
7. Haruno, M., Yamazaki, T.: High-performance bilingual text alignment using statistical and dictionary information. Natural Language Engineering **3**(1), 1–14 (1997)
8. Jean, S., Cho, K., Memisevic, R., Bengio, Y.: On using very large target vocabulary for neural machine translation. In: ACL 2015. pp. 1–10 (2015)
9. Kalchbrenner, N., Blunsom, P.: Recurrent continuous translation models. In: EMNLP 2013. pp. 1700–1709 (2013)

10. Lin, J., Sun, X., Ma, S., Su, Q.: Global encoding for abstractive summarization. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 2: Short Papers. pp. 163–169 (2018)
11. Lin, Z., Wang, X.: Chinese ancient-modern sentence alignment. In: Shi, Y., van Albada, G.D., Dongarra, J.J., Sloot, P.M.A. (eds.) International Conference on Computational Science (2). Lecture Notes in Computer Science, vol. 4488, pp. 1178–1185. Springer (2007)
12. Liu, Y., Wang, N.: Sentence alignment for ancient and modern Chinese parallel corpus. In: Lei, J., Wang, F.L., Deng, H., Miao, D. (eds.) Emerging Research in Artificial Intelligence and Computational Intelligence. pp. 408–415. Springer Berlin Heidelberg, Berlin, Heidelberg (2012)
13. Luong, T., Pham, H., Manning, C.D.: Effective approaches to attention-based neural machine translation. In: EMNLP 2015. pp. 1412–1421 (2015)
14. Luong, T., Pham, H., Manning, C.D.: Effective approaches to attention-based neural machine translation. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, EMNLP 2015, Lisbon, Portugal, September 17-21, 2015. pp. 1412–1421 (2015)
15. Ma, S., Sun, X., Wang, Y., Lin, J.: Bag-of-words as target for neural machine translation. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018, Melbourne, Australia, July 15-20, 2018, Volume 2: Short Papers. pp. 332–338 (2018)
16. Meng, F., Lu, Z., Li, H., Liu, Q.: Interactive attention for neural machine translation. In: COLING 2016. pp. 2174–2185 (2016)
17. Mi, H., Wang, Z., Ittycheriah, A.: Supervised attentions for neural machine translation. In: EMNLP 2016. pp. 2283–2288 (2016)
18. Resnik, P.: Parallel strands: A preliminary investigation into mining the web for bilingual text. In: Farwell, D., Gerber, L., Hovy, E.H. (eds.) Machine Translation and the Information Soup, Third Conference of the Association for Machine Translation in the Americas, AMTA '98, Langhorne, PA, USA, October 28-31, 1998, Proceedings. Lecture Notes in Computer Science, vol. 1529, pp. 72–82. Springer (1998). https://doi.org/10.1007/3-540-49478-2_7
19. Resnik, P.: Mining the web for bilingual text. In: Dale, R., Church, K.W. (eds.) ACL. ACL (1999)
20. See, A., Liu, P.J., Manning, C.D.: Get to the point: Summarization with pointer-generator networks. In: Barzilay, R., Kan, M. (eds.) Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 1: Long Papers. pp. 1073–1083. Association for Computational Linguistics (2017). https://doi.org/10.18653/v1/P17-1099
21. Su, J., Tan, Z., Xiong, D., Ji, R., Shi, X., Liu, Y.: Lattice-based recurrent neural network encoders for neural machine translation. In: Singh, S.P., Markovitch, S. (eds.) Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA. pp. 3302–3308. AAAI Press (2017)
22. Sutskever, I., Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks. In: Advances in neural information processing systems. pp. 3104–3112 (2014)
23. Sutskever, I., Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks. In: NIPS, 2014. pp. 3104–3112 (2014)
24. Tjandra, A., Sakti, S., Nakamura, S.: Local monotonic attention mechanism for end-to-end speech recognition. CoRR **abs/1705.08091** (2017)
25. Tu, Z., Lu, Z., Liu, Y., Liu, X., Li, H.: Modeling coverage for neural machine translation. In: ACL 2016 (2016)
26. Wang, X., Ren, F.: Chinese-japanese clause alignment. In: Gelbukh, A.F. (ed.) CICLing. Lecture Notes in Computer Science, vol. 3406, pp. 400–412. Springer (2005)