Modeling Coherence for Discourse Neural Machine Translation

Hao Xiong, Zhongjun He, Hua Wu, Haifeng Wang
Baidu Inc. No. 10, Shangdi 10th Street, Beijing, 100085, China
{xionghao05, hezhongjun, wu_hua, wanghaifeng} @baidu.com

Abstract

Discourse coherence plays an important role in the translation of one text. However, the previous reported models most focus on improving performance over individual sentence while ignoring cross-sentence links and dependencies, which affects the coherence of the text. In this paper, we propose to use discourse context and reward to refine the translation quality from the discourse perspective. In particular, we generate the translation of individual sentences at first. Next, we deliberate the preliminary produced translations, and train the model to learn the policy that produces discourse coherent text by a reward teacher. Practical results on multiple discourse test datasets indicate that our model significantly improves the translation quality over the state-of-the-art baseline system by +1.23 BLEU score. Moreover, our model generates more discourse coherent text and obtains +2.2 BLEU improvements when evaluated by discourse metrics.

Introduction

Discourse coherence, such as relevant conjunction for adjacent sentences, plays an important role in the translation of the text. However, standard Neural Machine Translation (NMT) models (Sutskever, Vinyals, and Le 2014; Bahdanau, Cho, and Bengio 2015) mainly focus on improving translation quality over individual sentence, and ignoring cross-sentence links and dependencies. With this preference, the translation of each sentence is independent of the other sentences, thus there is no guarantee to generate discourse coherent text.

Table 1 shows a concrete example from TED talks, where each sentence is rationally translated from the sentence perspective, but the missing conjunction 'And', and the missing coreference for the predicate 'build' cause the incoherence and poor readability of the entire text.

Intuitively, to generate better translation of the entire text, the model deserves considering the cross-sentence connections and dependencies, generating discourse coherent translations. Towards this demand, most previous work (Wang et al. 2017; Voita et al. 2018) proposed to explore additional context, generally is certain preceding adjacent sentences, to reinforce the model. However, the major goal of these models is still the quality of individual sentence while not the entire text, thus it is hard for them to generate promising discourse coherent translations.

Source	我们加入霓虹 我们加入柔和的粉蜡色 我们使用新型材料。 人们爱死这样的建筑了。 我们不断地建造。
Ref	We add neon and we add pastels and we use new materials. And you love it. And we can't give you enough of it.
NMT	We add the neon, we add soft, flexible crayons, and we use new materials. $[conj]_{miss}$People love architecture. $[conj]_{miss}$We keep building $[coref]_{miss}$.

Table 1: Instance of translation for one text consists of three sentences, where $[conj]_{miss}$ indicates missing cross-sentence conjunction, and $[coref]_{miss}$ indicates missing coreference. Although from the sentence perspective, the translations of second and third sentences produced by *NMT* system are acceptable, however the fluency of the entire text is poor.

To address this problem, we take an insight into human translation behavior for one text, where we first translate each sentence independently, and then take some modifications towards making the entire text coherently and fluently, such as replacing conjunctions and keeping the translation of terminologies consistently. Ideally, this procedure can be divided into two processes: 1) generating preliminary translation of each sentence, 2) deliberating each translation with the satisfaction of discourse coherence.

Motivated by the human translation behavior and the success of Deliberation Networks (Xia et al. 2017), we propose a two-pass decoder translation model, aiming at improving the coherence of the entire text. Specifically, we generate the preliminary translation of each sentence using the canonical NMT model, and then employ the Deliberation Networks to refine the translations over the entire text. However, since the standard NMT models focus on fine-tuning on local n-gram patterns, training by maximum likelihood estimation, the produced translations are generally locally coherent.

Intuitively, to generate discourse coherent translations, it requires training the model to learn the policy that receives more discourse coherent rewards straightforwardly. To satisfy this requirement, we explore a novel measure to estimate the quality of discourse coherence, namely a reward teacher (Bosselut et al. 2018), which learns the ordering structure in one text trained by a bidirectional recurrent neural networks (biRNN) (Schuster and Paliwal 1997). Rewarding by the reward teacher, the model learns to generate the discourse coherent translations while maintaining accurate translation for each individual sentence.

In particular, we design our model based on the Transformer architecture (Vaswani et al. 2017) according to its superior performance on machine translation tasks.

We evaluate the performance of our model on the IWSLT speech translation task with TED talks (Cettolo, Girardi, and Federico 2012) as training corpus, which includes multiple entire talks. Practical experiments reveal that our model improves the sentence translation quality by +1.23 BLEU (Papineni et al. 2002) score over one strong baseline. Moreover, when evaluated by discourse metrics, our model achieves an average improvements by +2.2 points in term of BLEU and +1.98 of METEOR (Denkowski and Lavie 2014). Through extensive experimental analysis, we confirm that our model can generate more discourse coherent translations than the baseline system.

To our knowledge, this is the first work on modeling coherence for discourse neural machine translation. The contributions of this paper can be concluded into the followings:

- We propose a two-pass decoder translation model to refine the translation of the discourse text.
- We propose a policy learning technique that encourages the model generating coherent and fluent discourse translations.
- We conduct extensive experiments to validate the effectiveness of our model, and analyze the results to confirm that our model can generate discourse coherent translations.

Our Approach

Intuitively, it is plausible that a model with external context summarized from entire text, trained by a discourse-aware reward can generate more accurate and discourse coherent translations. Towards exploring global context to refine the translation, we take inspirations from the work of Deliberation Network, to translate sentence in one text independently by the first-pass decoder, and then to summarize the first-pass translation as the external context for the second-pass decoder. Although most existing related work summarizes the external context with gold sequence, here we follow the original method of Deliberation Network that takes the predicted sequence as our global context, since it can also potentially alleviate the *exposure bias* problem (Ranzato et al. 2016).

To let the NMT model be prone to generate discourse coherent text, we learn the recent advances in text generation task (Bosselut et al. 2018), and propose to use the overall ordering structure of a document as an approximation of discourse coherent. As the two-pass decoders learn the policy that tends to receive more discourse rewards from the reward teacher, it is possible for the model satisfying the above mentioned requirements for the discourse translation.

Overall Architecture

In recent work, Vaswani et al.(2017) proposed an effective encoder-decoder based translation model, namely Transformer. The Transformer follows an encoder-decoder architecture using stacked self-attention and fully connected layers for both the encoder and decoder. In contrast to recurrent models, it avoids recurrence completely and drops the usage of complex Long Short-term Memory (LSTM) (Hochreiter and Schmidhuber 1997) and Gated Recurrent Unit (GRU) (Cho et al. 2014), being more parallelizable and faster to train. According to its effectiveness and superior performance on the translation task, we develop our architecture based on the Transformer model.

Figure 1 illustrates the overall architecture of our proposed two-pass decoder translation model. It is clear that our model can be divided into three important components: the first-pass decoder, the second-pass decoder and the discourse reward teacher. Since the original architecture of the Transformer is well designed, fine-tuning on the quality of sentence translation, we replicate the original Transformer encoder-decoder architecture in the first-pass decoder.

As mentioned in the literature (Voita et al. 2018), techniques developed for the recurrent based models did not present effective for the Transformer, instead they proposed two separated encoders to learn the representations for the context and source sequence respectively. Notably, the context used in their model is one preceding adjacent sentence appeared in the document. However, after preliminary experiments, we found that mechanically replication of previous techniques yields bad performance in our architecture. Instead we stack an additional self-attention layer in the canonical Transformer decoder, resulting in three types of self-attention layers in the second-pass decoder. Ideally, the additional self-attention layer will summarize the context from the translation of overall document generated by the first-pass decoder, making the decoder to generate discourse coherent translation as possible, since it has learned the potential translations of other sentences in this document.

The last important component of our model is the discourse reward teacher, an offline trained biRNNs that rewards the model to generate more discourse coherent translation. Before we describe the discourse reward teacher in details, we firstly draw some definitions of our two-pass decoder for better understanding of this paper.

First-pass Decoder

We define a source document of n sentences as $S_x = \{s_{x:0}, ..., s_{x:n-1}\}$ where each sentence $s_{x:i}$ have T_i words. The goal of first-pass decoder is trained to minimize the negative log-likelihood of predicting the target word, y_t:

$$L_{mle1} = -\sum_{i}^{n}\sum_{t}^{T_i} \log P(y_t|y_0, ..., y_{t-1}, \mathrm{H}_{enc}, \mathrm{H}_{dec1}) \quad (1)$$

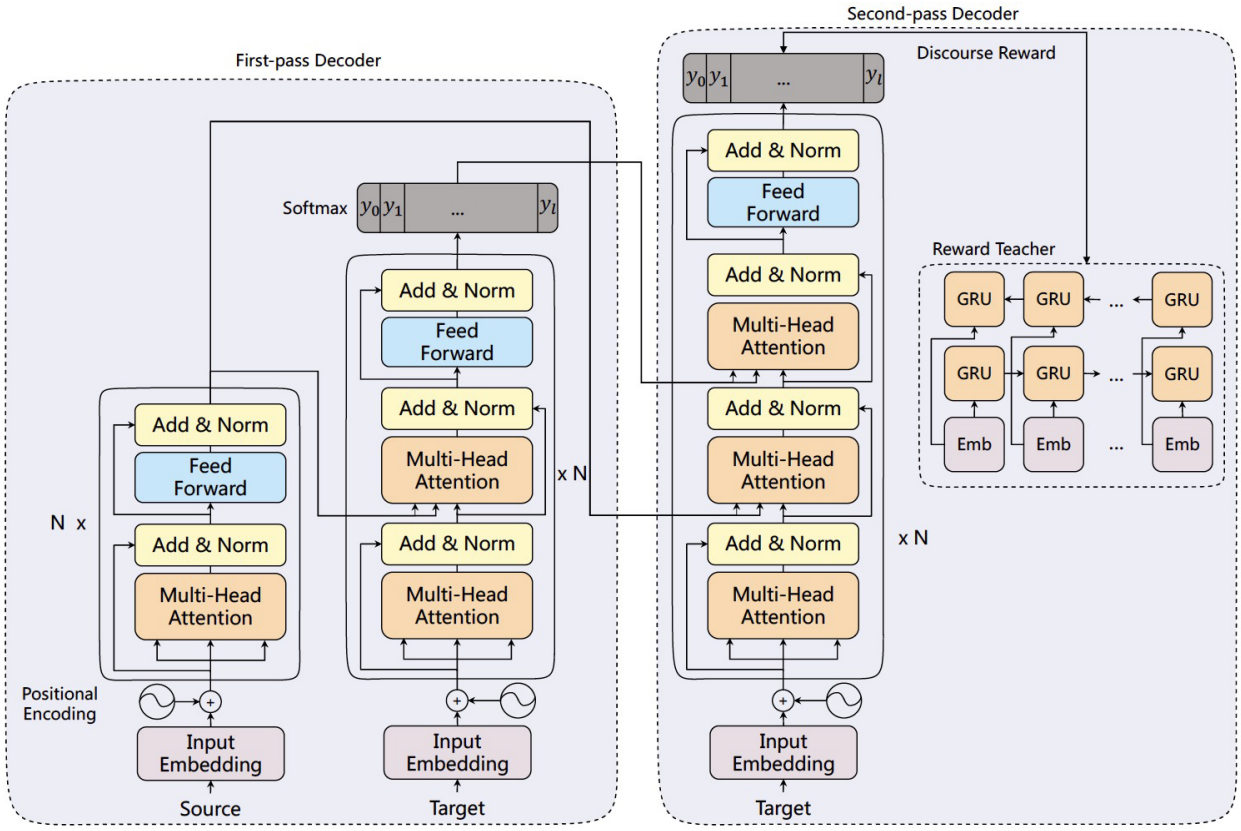

Figure 1: Illustration of the overall architecture of our two-pass decoder translation model. The first-pass decoder produces translations as the canonical Transformer model does. And the second-pass decoder utilizes an additional self-attention layer, exploring more contextual information from the other sentences generated by the first-pass decoder. We also let the model learning the policy to generate more fluent and coherent translation by rewarding from a reward teacher.

where T_i is the length of the target sequence, and the H_{enc}, H_{dec1} are the representations of the encoder and the decoder in the first-pass decoder.

Second-pass Decoder

In contrast to the original Deliberation Network, where they proposed a complex joint learning framework to train the model. By virtue of our preliminary experiments, we found that a simpler solution is enough to obtain promising results. Thus, we treat the first-pass decoder and the second-pass decoder as two associative learning tasks by sharing the identical encoder, minimizing the following loss:

$$L_{mle} = L_{mle1} + L_{mle2} \tag{2}$$

where

$$L_{mle2} = -\sum_{i}^{n}\sum_{t}^{T_i} \log P(y_t|y_0, ..., y_{t-1}, \mathrm{Y}^{'}, \mathrm{H}_{enc}, \mathrm{H}_{dec2}) \tag{3}$$

where H_{dec2} is the representation of the decoder in the second-pass decoder. Here we use different training parameters for decoders in the first-pass and the second-pass decoding stage independently. As in the first-pass decoder, we can use a greedy strategy or beam search to generate the preliminary document translations, $\mathrm{Y}^{'}$. Generally, $\mathrm{Y}^{'}$ is represented by word embeddings.

Reward Teacher

In recent work, Bosselut et al.(2018) proposed two neural teachers, absolute order teacher and relative order teacher to model the coherence of generated text towards the text generation task. Motivated by their success, we extend this approach to the machine translation task, since in our task we need also generate coherent translations which is the same as in the text generation task.

Specifically, we develop absolute order teacher as our reward teacher by training a sentence encoder to minimize the similarity between a sequence encoded in its forward order, and the same sequence encoded in the reverse order. [1] Following the same definitions as Bosselut et al.(2018), each

[1] Although Bosselut et al.(2018) obtains better results with the relative order teacher in their experiments, we found in the machine translation task an absolute order teacher is appropriate to obtain promising performance and omits the descriptions of the relative order teacher.

target sentence $s_{y:i}$ that has L_i words, is represented as:

$$s_{y:i} = \sum_{j}^{L_i} y_{ij} \tag{4}$$

where y_{ij} is a word embedding and $s_{y:i}$ is a sentence embedding.

Identical to the canonical encoder in the recurrent models, each $s_{y:i}$ is passed to a GRU:

$$h_i = \mathrm{GRU}(h_{i-1}, s_{y:i}) \tag{5}$$

and the final hidden state of the RNN is utilized as the representation of the document:

$$f(\mathrm{S_y}) = h_n \tag{6}$$

where $f(\mathrm{S_y})$ is the representation of the document and h_n is the final state of the recurrent neural networks.

Intuitively, if one text is well organized, the similarity between the sentence embedding from reading the sentences in the forward order, and from reading the sentences in the reverse order, should be minimized.

Thus, the absolute order teacher is trained to minimize the cosine similarity between two orders:

$$L_{abs} = \frac{\langle f(\overrightarrow{\mathrm{S_y}}), f(\overleftarrow{\mathrm{S_y}})\rangle}{\|f(\overrightarrow{\mathrm{S_y}})\|\|f(\overleftarrow{\mathrm{S_y}})\|} \tag{7}$$

After training on the monolingual corpus, we use this learned teacher to generate a reward that judges the generated sequence's ordering similarity to the gold sequence.

Notably, the reward teacher is trained offline on gold sequences in an unsupervised manner prior to training the NMT model, and its parameters are fixed during policy learning.

Policy Learning

Since training the two-pass decoder with maximum likelihood estimation produces translations that are locally coherent, so there is no guarantee to generate coherent discourse. As mentioned in the above section, we train a reward teacher to reward the model that generates good ordering structure, encouraging the model to learn a policy that produces discourse coherent translation explicitly. In this paper, we learn a policy using the self-critical training approach (Rennie et al. 2017).

Specifically, for each training example (one document), a sequence $\hat{y}$ is generated by sampling from the models's distribution $P(\hat{y}_t|\hat{y}_0, ..., \hat{y}_{t-1}, \mathrm{Y}^{'}, \mathrm{H}_{enc}, \mathrm{H}_{dec2})$ of the second-pass decoder. Another sequence y^* is generated by argmax decoding from $P(y^*_t|y^*_0, ..., y^*_{t-1}, \mathrm{Y}^{'}, \mathrm{H}_{enc}, \mathrm{H}_{dec2})$ at each time step t.

The model is trained to minimize:

$$\begin{aligned} L_{rl2} &= -\sum_{i}^{n}\sum_{t}^{T_i} R \cdot \log P(y_t|y_0, ..., y_{t-1}, \mathrm{Y}^{'}, \mathrm{H}_{enc}, \mathrm{H}_{dec2}) \\ R &= r(\hat{y}) - r(y^*) \end{aligned} \tag{8}$$

where $r(y^*)$ is the reward produced by the reward teacher for the greedily decoded sequence, and $r(\hat{y})$ is the reward for the sampled sequence.

Since $r(y^*)$ can be viewed as a baseline reward, the model learns to generate sequence that receives more reward from the teacher than the best sequence, which can be greedily decoded from the current policy. This approach allows the model to explore sequence that yields higher reward than the current best policy.

Notably, here we introduce the policy learning for the second-pass decoder. In actual, it is natural facilitating the first-pass decoder with the identical technique, which will be described later.

Absolute Order Reward

Once the sequences $\hat{y}$, y^* are generated, we use the absolute order reward teacher to reward these sequences:

$$\begin{aligned} r(\hat{y}) &= \frac{\langle f(\overrightarrow{\mathrm{S_{\hat{y}}}}), f(\overrightarrow{\mathrm{S_y}})\rangle}{\|f(\overrightarrow{\mathrm{S_{\hat{y}}}})\|\|f(\overrightarrow{\mathrm{S_y}})\|} - \frac{\langle f(\overrightarrow{\mathrm{S_{\hat{y}}}}), f(\overleftarrow{\mathrm{S_y}})\rangle}{\|f(\overrightarrow{\mathrm{S_{\hat{y}}}})\|\|f(\overleftarrow{\mathrm{S_y}})\|} \\ r(y^*) &= \frac{\langle f(\overrightarrow{\mathrm{S_{y^*}}}), f(\overrightarrow{\mathrm{S_y}})\rangle}{\|f(\overrightarrow{\mathrm{S_{y^*}}})\|\|f(\overrightarrow{\mathrm{S_y}})\|} - \frac{\langle f(\overrightarrow{\mathrm{S_{y^*}}}), f(\overleftarrow{\mathrm{S_y}})\rangle}{\|f(\overrightarrow{\mathrm{S_{y^*}}})\|\|f(\overleftarrow{\mathrm{S_y}})\|} \end{aligned} \tag{9}$$

where $f(\overrightarrow{\mathrm{S_y}})$ is the representation of forward-ordered corresponding gold sequence and $f(\overleftarrow{\mathrm{S_y}})$ is the representation of reverse-ordered gold sequence.

This reward compares the generated sequence to both sentence orders of the gold sequence, and rewards sequences that are more similar to the forward order of the gold sequence. Because the cosine similarity terms in Equation (9) are bounded in $[-1; 1]$, the model receives additional reward for generating sequences that are different from the reverse-ordered gold sequence.

Joint Learning

There are two decoders in our model, the first-pass decoder and the second-pass decoder, each of them can learn parameters to minimize the negative log-likelihood independently. Intuitively, these two decoders are associative, and both performance can be improved by the joint learning techniques as shown in Equation (2).

As aforementioned, we use a reward teacher to reward the model generating discourse coherent text. According to our architecture, there are two approaches to reward the model learning the policy. One is described in the previous section that rewards the second-pass decoder by the self-critical learning strategy. We argue that the performance can be further improved when the first-pass decoder is also rewarded by the reward teacher. Thus, the final objective of our model is to minimize:

$$L = L_{mle1} \cdot \lambda_1 + L_{rl1} \cdot (1 - \lambda_1) + L_{mle2} \cdot \lambda_2 + L_{rl2} \cdot (1 - \lambda_2) \tag{10}$$

where λ_1 and λ_2 are two hyperparameters that balance learning the discourse-focused policy while maintaining the accurate translation.

The losses, L_{mle1}, L_{mle2}, L_{rl2} are introduced in the Equation (1), Equation (3) and Equation (8) respectively. The computation of L_{rl1} is almost identical to the L_{rl2} with

slight modification by replacing the model' distribution from the first-pass decoder:

$$L_{rl1} = -\sum_{i}^{n}\sum_{t}^{T_i} R \cdot \log P(y_t|y_0, ..., y_{t-1}, \mathrm{H}_{enc}, \mathrm{H}_{dec1})$$
$$R = r(\hat{y}) - r(y^*) \tag{11}$$

where $r(\hat{y})$ and $r(y^*)$ can be computed by the Equation (9).

Experiments

We evaluate our model on the IWSLT 2015 Chinese-English translation task with TED talks as our training corpus, since it includes entire discourse text.

Data Preprocess

Considering the memory capacity, we split one talk that has more than 16 sentences into several small talks, ensuring the experiments can be successfully conducted on most GPUs.

Specifically, we take the *dev-2010* as our development set, and *tst-2013∼2015* as our test sets. Statistically, we have 14,258 talks and 231,266 sentences in the training data, 48 talks and 879 sentences in the development set, and 234 talks and 3,874 sentences in the test sets.

Following the work of Chao and Zong2017 , we conduct byte-pair encoding (Sennrich, Haddow, and Birch 2016) for both Chinese and English sentences, setting the vocabulary size to 20K and 18K respectively.

For English tokenization, we use the script supplied by Moses Toolkit [2]. And we segment the Chinese sentences into words by an open source toolkit [3].

Systems

We measure the performance of our model with different system implementations.

- *t2t*: This is the official supplied open source toolkit for running Transformer model. Specifically, we use the v1.6.5 release [4].
- *context-encoder*: The reimplementation of the work Voita et al.(2018).
- *first-pass*: This system is applied to minimize the Equation (1). Actually, this system is almost identical to the *t2t* but with different data shuffling strategy, where the *t2t* shuffles the data over the sentences randomly, while the *first-pass* shuffles the data over the talks.
- *first-pass-rl*: We implement this system to minimize $L_1 = L_{mle1} \cdot \lambda_1 + L_{rl1} \cdot (1 - \lambda_1)$, which is the first part of the Equation (10). This system is to evaluate the policy learning strategy for the standard Transformer model.
- *two-pass*: This system is applied to minimize the Equation (2), to measure the contribution of the two-pass decoder strategy.
- *two-pass-rl*: This is the final system, tuning to minimize the Equation (10).

We also implement the reward teacher with standard biRNNs, minimizing the Equation (7), namely *reward-teacher*.

Training Details

Since the size of training data is relatively small in the NMT task, we use the *base* version hyperparameters of the standard Transformer model, against model overfitting. For all systems, we use the Adam Optimizer (Kingma and Ba 2015) with the identical settings to $t2t$, to tune the parameters.

One thing deserves to be noted is the value of hyperparameter *batch_size*. In general, a large value of batch size achieves better performance when training on large scale corpus (more than millions) (Vaswani et al. 2017). However, in our preliminary experiments we found that a smaller value presented better results training on the TED talks. Thus we set the *batch_size* to 320 for *t2t* system, resulting in approximately 10∼20 sentences in one batch according to the different sentence length.

For the other systems, we read one talk per one batch, producing no more than 16 sentences in one batch, which is comparable to the baseline system, *t2t*.

Following the work of Bosselut et al.(2018), we train the *reward-teacher* with the similar hyperparmaters but using different optimizing strategy. Specifically, we set both the embedding and recurrent hidden size to 100, and apply one dropout layer with keeping probability equals to 0.3 between the embedding layer and the bidirectional recurrent layers. For tuning the parameters, we use the same Adam Optimizer as the NMT systems.

Notably, we train the *reward-teacher* with the monolingual English from the TED talks as baseline system, and investigate the effect on the translation quality when the *reward-teacher* trained with variant monolingual datas in later section.

The training speed of *two-pass-bleu-rl* model is 8 talks per one second running on V100 with 8GPUs, and it needs about 1.5 days to converge.

Results and Analysis

To measure the performance of our systems, we use the universal BLEU and METEOR metrics, computed by two open source toolkits [5] [6]. In addition to measuring the quality for each individual sentence, we also concatenate sentences in one talk into one long sentence, and then represent its BLEU and METEOR scores as BLEU_{doc}, and METEOR_{doc}.

[2] https://github.com/moses-smt/mosesdecoder/blob/master/scripts/tokenizer/tokenizer.perl

[3] https://github.com/fxsjy/jieba

[4] https://github.com/tensorflow/tensor2tensor/releases/tag/v1.6.5

[5] https://github.com/moses-smt/mosesdecoder/blob/master/scripts/generic/multi-bleu.perl

[6] http://www.cs.cmu.edu/~alavie/METEOR/index.html#Download

SYSTEMS	BLEU	METEOR	BLEU_{doc}	METEOR_{doc}
t2t (Vaswani et al. 2017)	20.10	35.75	25.60	34.91
context-encoder (Voita et al. 2018)	20.31	35.79	25.93	35.03
first-pass	20.41	35.98	25.99	35.23
first-pass-rl	20.79	36.41	26.82	36.12
two-pass	20.92	36.70	26.89	36.29
two-pass-rl	21.11	36.82	27.50	36.64
two-pass-bleu	21.08	36.88	27.32	36.42
two-pass-bleu-rl	**21.33**	**36.94**	**27.8**	**36.89**
Pretrain with 25M bilingual training corpus from WMT2018				
t2t	26.57	42.80	31.65	39.47
two-pass-bleu-rl	**27.55**	**43.77**	**33.98**	**41.28**

Table 2: Performance of systems measured by different metrics. Due to the space limitation, we list an average score of all test datasets. To measure the discourse quality, we concatenate sentences in one talk into one long sentence, and then evaluate their BLEU and METEOR scores as BLEU_{doc} and METEOR_{doc}.

Overall Results From the Table 2, it is interesting that our *first-pass* system beats the baseline *t2t* slightly. As we described in the previous section, the difference between such two systems is the diverse shuffling strategy. Different from the *t2t* system, where we shuffle the overall training data and select certain sentences into one batch randomly, while in the *first-pass* system, we shuffle the talks and take sentences from one talk into one batch orderly. This finding indicates that the discourse text with ordering structure is better trained with its original order while not to be scattered, ensuring the effectiveness of our other systems since they are all trained like the *first-pass* system. In addition, this training strategy can be viewed as well designed curriculum learning (Bengio et al. 2009) strategy, which has been proved effective for NMT task (Kocmi and Bojar 2017).

Compared to the *first-pass* system, our *two-pass* system performs better results on these test datasets, which confirms that using deliberation procedure can improve the translation quality, although we implement a slightly different Deliberation Networks. When evaluated by the discourse metrics, we find that the Deliberation Networks is able to generate discourse coherent translation, since it can bring more average improvements compared to the sentence metrics (+0.51 vs +0.8 in term of BLEU, and +0.72 vs +1.06 in term of METEOR).

When examining models trained using a reward teacher, the systems (*first-pass-rl* and *two-pass-rl*) achieve significant improvements over these systems trained by the standard maximum likelihood estimation (*first-pass* and *two-pass*), by means of +0.72 BLEU_{doc} and +0.62 METEOR_{doc} when evaluated by the discourse metrics. Moreover, our systems can generate discourse coherent text while maintaining improving quality on each sentence, with the evidence that two systems also improve the translation quality at sentence-level when evaluated by the sentence metrics (+0.29 BLEU, +0.28 METEOR).

Another finding is that when given more context (up to 16 surrounding sentences), the *two-pass* system achieves more improvements compared to the *context-encoder* system which models one preceding sentence as external context. It deserves researching in the future that exploring more context to improve the translation quality.

As shown in the last row, when we take the discourse BLEU score as additional reward, our model beats the baseline system by +1.23 BLEU and +1.19 METEOR. When evaluated by discourse metrics, the improvement is more significant, by +2.2 BLEU and +1.98 METEOR, proving the effectiveness of our model.

Pretrain We also investigate using a large-scale corpus, Chinese-English training corpus from the WMT2018 translation task, to pretrain the model, and then fine-tune on the TED corpus. For *two-pass-bleu-rl* system, we pretrain all parameters except one special self-attention layer, which is responsible for capturing relationships between current sentence and first-pass produced coarse translations.

From the Table 2, it is clear that pretrained models obtain more than 6 points improvements. Also, after pretraining, our *two-pass-bleu-rl* system still obtains improvements upon the baseline system, which indicates that our approach is robust and practical.

Effect of Balance Factor We investigate the effect of two balance factors (λ_1 and λ_2 in Equation (10)) on the performance of the translation quality evaluated by different type of metrics.

We first adjust the value of λ_1, ranging from 0.7 to 1.0[7], and stepping by 0.05, to see the performance change for *first-pass-rl* system. Next, we then adjust the value of λ_2 with fixed value of λ_1, to optimize the performance of *two-pass-rl* system.

As shown in Figure 2, we see that setting the value of λ_1 to 0.85 and λ_2 to 0.80 produces the best performance for *first-pass-rl* and *two-pass-rl*.

Another finding is that the change of sentence-level BLEU score appears to be slightly consistent with the discourse BLEU_{doc} score, but the latter has larger fluctuation (two top lines), indicating the reward teacher encourages generating discourse coherent translation explicitly.

[7] According to our preliminary experiments, we find that a value lower than 0.7 failed to produce reasonable translations.

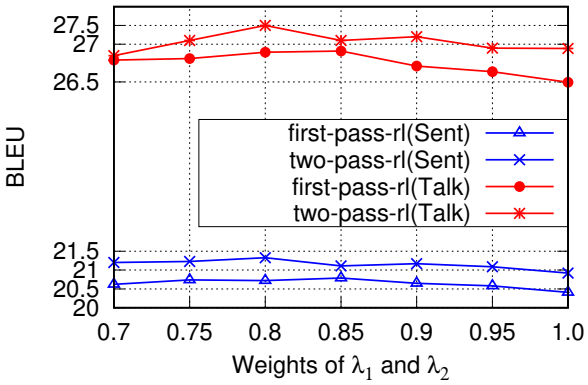

Figure 2: Effect of two balance factors on the performance of the translation quality.

Effect of Reward Teacher Since the parameters for the *reward-teacher* are fixed during the translation stage, we can explore variant reward teachers trained by different corpus, and see the effect on the translation performance.

METRICS	TED	AFP	XIN	Gigaword
BLEU	21.11	20.54	20.92	20.49
BLEU_{doc}	27.50	26.77	27.32	26.83

Table 3: Performance of *two-pass-rl* rewarded by variant reward teachers that trained by different corpus. AFP and XIN are the corresponding proportion of English Gigaword.

Except the TED corpus, we train the *reward-teacher* with English Gigaword Fifth Edition [8] that includes larger English discourse text but was written in different style. As shown in Table 3, the teacher trained by external corpus yields bad performance, despite its large corpus size. This suggests that the reward teacher deserves to be trained with the corpus that was written in the similar style.

Coherence Lapata and Barzilay(2005) proposed one approach that measures discourse coherence as sentence similarity. Specifically, the representation of each sentence is the mean of the distributed vectors of its words, and the similarity between two sentences S_1 and S_2, is determined by the cosine of their means:

$$sim(S_1, S_2) = \frac{\langle f(S_1), f(S_2) \rangle}{\|f(S_1)\| \|f(S_2)\|} \quad (12)$$

where $f(S_i) = \sum_{w \in S_i} \vec{w}$, and $\vec{w}$ is the vector for word w.

We use Word2Vec [9] to learn the distributed vectors of words by training on the aforementioned English Gigaword Fifth Edition. And we set the dimensionality of word embeddings to 100.

[8] https://catalog.ldc.upenn.edu/LDC2011T07

[9] http://word2vec.googlecode.com/svn/trunk/

SYSTEMS	*tst-2013*	*tst-2014*	*tst-2015*
t2t	0.5991	0.5838	0.5939
first-pass	0.5999	0.5845	0.5943
two-pass	0.6011	0.5880	0.5962
first-pass-rl	0.6008	0.5861	0.5952
two-pass-rl	0.6032	0.5913	0.6008
two-pass-bleu-rl	0.6041	0.5938	0.6014
human translation	0.6066	0.5910	0.6013

Table 4: We measure the discourse coherence as sentence similarity.

Table 4 shows the cosine similarity of adjacent sentences on all test datasets. It reveals that systems encouraged by discourse reward produce better coherence in document translation than contrastive systems in term of cosine similarity.

Conjunctions We count the top five frequent conjunctions in the translations produced by *t2t* and *two-pass-rl*, to see the concrete transformation of sentences that encouraged to generate coherent translations.

t2t	And (519) But (186) In (114) So (174) What (55)
*sys**	And (540) But (183) In (129) So (178) What (73)

Table 5: The statistics of top five frequent conjunctions in two systems (*sys** is *two-pass-bleu-rl*). Numbers in bracket is the occurrences of this word.

As shown in Table 5, sentences in *two-pass-bleu-rl* tend to using more diverse conjunctions to build the connections towards preceding sentence, which proves that our model can generate more discourse coherent translation.

Related Work

Gong, Zhang, and Zhou(2011) proposed a memory based approach to capture contextual information to facilitate the statistical translation model generating discourse coherent translations, and the literatures (Kuang et al. 2017; Tu et al. 2018; Maruf and Haffari 2018) extended similar memory based approach to the NMT framework.

Wang et al.(2017) presented a novel document RNN to learn the representation of the entire text, and treated the external context as the auxiliary context which will be retrieved by the hidden state in the decoder.

Tiedemann and Scherrer(2017) and Voita et al.(2018) proposed to encode global context through extending the current sentence with one preceding adjacent sentence. Notably, the former was conducted on the recurrent based models while the latter was implemented on the Transformer model.

Conclusion and Future Work

In this paper, we propose two novel techniques, Deliberation Networks and reward teacher to generate discourse coherent translation. Practical experiments confirm, through modeling external discourse context from the potential translations of the other sentences in the same text, our model can

improve the translation quality both on the sentence-level (+1.23 BLEU) and discourse-level (+2.2 BLEU) metrics. Moreover, when the model learns the policy that rewarded by a reward teacher, it can generate more fluent and coherent discourse translations. In the future, we will continue research on using more bilingual training data that has no explicit discourse boundaries, and verify our model on multilingual translation tasks.

References

Bahdanau, D.; Cho, K.; and Bengio, Y. 2015. Neural machine translation by jointly learning to align and translate. In *International Conference on Learning Representations*.

Bengio, Y.; Louradour, J.; Collobert, R.; and Weston, J. 2009. Curriculum learning. In *Proceedings of the 26th annual international conference on machine learning*, 41–48. ACM.

Bosselut, A.; Celikyilmaz, A.; He, X.; Gao, J.; Huang, P.-S.; and Choi, Y. 2018. Discourse-aware neural rewards for coherent text generation. In *Proceedings of NAACL-HLT 2018*, 173–184. Association for Computational Linguistics.

Cettolo, M.; Girardi, C.; and Federico, M. 2012. Wit^3: Web inventory of transcribed and translated talks. In *Proceedings of the 16th Conference of the European Association for Machine Translation (EAMT)*, 261–268.

Chao, B., and Zong, H. 2017. Towards better translation performance on spoken language. In *International Workshop on Spoken Language Translation (IWSLT)*.

Cho, K.; van Merrienboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; and Bengio, Y. 2014. Learning phrase representations using rnn encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1724–1734.

Denkowski, M., and Lavie, A. 2014. Meteor universal: Language specific translation evaluation for any target language. In *Proceedings of the EACL 2014 Workshop on Statistical Machine Translation*.

Gong, Z.; Zhang, M.; and Zhou, G. 2011. Cache-based document-level statistical machine translation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 909–919. Association for Computational Linguistics.

Hochreiter, S., and Schmidhuber, J. 1997. Long short-term memory. *Neural computation* 9(8):1735–1780.

Kingma, D. P., and Ba, J. 2015. Adam: A method for stochastic optimization. In *International Conference for Learning Representation (ICLR)*.

Kocmi, T., and Bojar, O. 2017. Curriculum learning and minibatch bucketing in neural machine translation. In *Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017*, 379–386. INCOMA Ltd.

Kuang, S.; Xiong, D.; Luo, W.; and Zhou, G. 2017. Cache-based document-level neural machine translation. *arXiv preprint arXiv:1711.11221*.

Lapata, M., and Barzilay, R. 2005. Automatic evaluation of text coherence: Models and representations. In *IJCAI*, volume 5, 1085–1090.

Maruf, S., and Haffari, G. 2018. Document context neural machine translation with memory networks. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1275–1284. Association for Computational Linguistics.

Papineni, K.; Roukos, S.; Ward, T.; and Zhu, W.-J. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting on association for computational linguistics*, 311–318. Association for Computational Linguistics.

Ranzato, M.; Chopra, S.; Auli, M.; and Zaremba, W. 2016. Sequence level training with recurrent neural networks. In *International Conference on Learning Representations*.

Rennie, S. J.; Marcheret, E.; Mroueh, Y.; Ross, J.; and Goel, V. 2017. Self-critical sequence training for image captioning. In *CVPR*, volume 1, 3.

Schuster, M., and Paliwal, K. K. 1997. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing* 45(11):2673–2681.

Sennrich, R.; Haddow, B.; and Birch, A. 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)*, volume 1, 1715–1725.

Sutskever, I.; Vinyals, O.; and Le, Q. V. 2014. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, 3104–3112.

Tiedemann, J., and Scherrer, Y. 2017. Neural machine translation with extended context. In *Proceedings of the Third Workshop on Discourse in Machine Translation*, 82–92.

Tu, Z.; Liu, Y.; Shi, S.; and Zhang, T. 2018. Learning to remember translation history with a continuous cache. *Transactions of the Association of Computational Linguistics* 6:407–420.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*, 5998–6008.

Voita, E.; Serdyukov, P.; Sennrich, R.; and Titov, I. 2018. Context-aware neural machine translation learns anaphora resolution. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1264–1274. Association for Computational Linguistics.

Wang, L.; Tu, Z.; Way, A.; and Liu, Q. 2017. Exploiting cross-sentence context for neural machine translation. In *Conference on Empirical Methods in Natural Language Processing*.

Xia, Y.; Tian, F.; Wu, L.; Lin, J.; Qin, T.; Yu, N.; and Liu, T.-Y. 2017. Deliberation networks: Sequence generation beyond one-pass decoding. In *Advances in Neural Information Processing Systems*, 1784–1794.