

A Implementation

Network Architecture. The model architecture of generator and reconstructor is almost the same except the length of input and output sequence. We adapt model architecture for our generator and reconstructor from (See et al., 2017) who used hybrid-pointer network for text summarization. The hybrid-pointer networks of generator and reconstructor are all composed of two one-layer unidirectional LSTMs as its encoder and decoder, respectively, with a hidden layer size of 600. Since we use two kinds of methods on adversarial training, there are two discriminators with different model architecture. In the Section 5.1, the discriminator is composed of four residual blocks with 512 hidden dimensions. While in Section 5.2, we use only one layer unidirectional LSTM with a hidden size of 512 as our discriminator.

Details of Training. In all experiments except in Section 6.4, we set the weight α in (2) controlling R_{loss} to 25. In Section 6.4, to prevent generator from overfitting to sentences from CNN/Daily Mail summary, we set the weight α to 50 which was larger than other experiments. We find that the if the value of α is too large, generator will start to generate output unlike human-written sentences. On the other hand, if the value of α is too small, the sentences generated by generator will sometimes become unrelated to input text of generator. For all the experiments, the baseline b in (3) gradually decreases from 0.25 to zero within 10000 updates on generator.

We set the weight β_1 of the gradient penalty in Section 5.1 to 10, and used RMSPropOptimizer with a learning rate of 0.00001 and 0.001 on the generator and discriminator, respectively. In Section 5.2.1, the weight β_2 of gradient penalty terms was 1.0, and used RMSPropOptimizer with a learning rate of 0.00001 and 0.001 on the generator and discriminator, respectively. It's also feasible to apply weight clipping in discriminator training, but the performance of gradient penalty trick was better.

B Corpus Pre-processing

- English Gigaword: We used the script of (Rush et al., 2015) to construct our training and testing datasets. The vocabulary size was set to 15K in all experiments.
- CNN/Diary Mail: We obtained 287227 train-

ing pairs, 13368 validation pairs and 11490 testing pairs identical to (See et al., 2017) by using the scripts provided by (See et al., 2017). To make our model easier to train, during training and testing time, we truncated input articles to 250 tokens (original articles has 781 tokens on average) and restricted the length of generator output summaries (original summaries has 56 tokens on average) to 50 tokens. The vocabulary size was set to $15k$.

- Chinese Gigaword: The Chinese Gigaword is a long text summarization dataset which is composed of 2.2M paired data of headlines and news. We preprocessed the raw data as following. First, we selected the 4K most frequent Chinese characters to form our vocabulary. We filtered out headline-news pairs with excessively long or short news segments, or that contained too many out-of-vocabulary Chinese characters, yielding 1.1M headline-news pairs from which we randomly selected 5K headline-news pairs as our testing set, 5K headline-news pairs as our validation set, and the remaining pairs as our training set. During training and testing, the generator took the first 80 Chinese characters of the source text as input.

C Model Pre-training

As we found that the different pre-training methods for the generator influenced final performance dramatically in all of the experiments, we felt it was important to find a proper unsupervised pre-training method to help the machine grasp semantic meaning. The summarization tasks on two datasets is different: One is sentence summarization, while the other is long text summarization. Therefore, we used the different pre-training strategies on two datasets described below.

- CNN/Diary Mail: The CNN/Diary Mail is a long text summarization dataset in which the source text consists of several sentences. Given the previous $i - 1$ sentences $sent_0, sent_1, \dots, sent_{i-1}$ from the source text, the generator predicted the next four sentences $sent_i, sent_{i+1}, \dots, sent_{i+3}$ in the source text as its pre-training target. If more than 40% of the words in target sentences $sent_i, sent_{i+1}, \dots, sent_{i+3}$ did not appear in

the given text, we filtered out this pre-training sample pair. This pre-training method allowed the generator to capture the important semantic meanings of the source text. Although the first few sentences of articles in CNN/Diary Mail contains the main information of articles, we hope we can provide a more general pre-training method which don't have any assumption of dataset and can be easily applied to other datasets.

- Chinese Gigaword: The pre-training method of Chinese Gigaword was similar to CNN/Diary Mail except that generator predicted the next sentence instead of next consecutive four sentences.
- English/Chinese Gigaword: As the source text of English Gigaword is made up of only one sentence, it is not feasible to split the last sentence from the source text; hence the previous pre-training method on Chinese Gigaword is not appropriate for this dataset. To properly initialize the set, we randomly selected 6 to 11 consecutive words in the source text, after which we randomly swapped 70% of the words in the source text. Given text with incorrect word arrangements, the generator predicted the selected words in the correct arrangement. We pre-trained in this way because we expect the generator to initialize with a rough language model. In Chinese Gigaword we also conducted experiments on pre-training in this manner, but the results were not as good as those shown in the part (C) of Table 3. In addition, we also used the retrieved paired data in row (B-1) in Table 1 to pre-train the generator in English Gigaword. However, pre-training generator with this method doesn't yield results better than those in Table 1.
- Transfer Learning: Before unsupervised training, the generator was pre-trained with paralleled data on CNN/Daily Mail dataset. However, the characteristics for two datasets are different. In English Gigaword, the articles were short and the summaries consist of only one sentence, while in CNN/Daily Mail dataset, the articles were extremely long and summaries consist of several sentences. To overcome these differences, during pre-training time, we took the first 35-45 words

in each CNN/Diary Mail article as generator input, and generator randomly predicted one of the sentences of the article's summary. In addition, we used the one sentence from CNN/Diary Mail summaries as real data to discriminator instead full summaries.

D Examples

Source Text: south korea issued a stern warning monday against illegal labor disputes and campus protests and announced the arrest of ### radicals for violent weekend disturbances .	
Ground Truth: south korea issues stern warning against labor and campus activists	(A-1)Supervised Result: south korea issues stern warning against illegal labor disputes
(C-2)WGAN: south korea issued stern warning against illegal labor disputes	(C-3)Adversarial REINFORCE: south korea issued stern warning against illegal labor disputes campus arrest
(E-2)WGAN : south korea issued stern warning against illegal labor disputes and arrest	(E-3)Adversarial REINFORCE: south korea issued stern warning against illegal labor disputes campus protests

Figure 5

Source Text: the thai government will set up an independent committee to investigate the management and lending practices of the state-owned krung thai bank -lrb- ktb -rrb- , channel # , thailand 's state television station reported monday .	
Ground Truth: thailand to investigate state-owned bank	(A-1)Supervised Result: thai government to set up committee to investigate lending practices
(C-2)WGAN: thai set up independent committee to investigate management	(C-3)Adversarial REINFORCE: thai government set up independent committee to investigate management lending practices
(E-2)WGAN : the thai committee will set up an independent committee to investigate the management	(E-3)Adversarial REINFORCE: thai government will set up an independent committee to investigate

Figure 8

Source Text: dutch police arrested monday ## environmentalists who were attempting to prevent the destruction of a <unk> -lrb- ##-acre -rrb- woods next to an airport used by nato forces , the anp news agency reported .	
Ground Truth: dutch police arrest ## environmental protesters	(A-1)Supervised Result: dutch police arrest ## environmentalists over nato destruction
(C-2)WGAN: dutch police arrested ## environmentalists attempting to prevent destruction	(C-3)Adversarial REINFORCE: dutch police arrested ## environmentalists attempting to prevent destruction of woods
(E-2)WGAN : dutch police arrested who were attempting to prevent destruction of a -lrb-	(E-3)Adversarial REINFORCE: dutch police arrested monday ## environmentalists who were attempting to prevent destruction woods

Figure 6

Source Text: the bewildering fight between the government and telemarketers over the national do-not-call list took another turn when a second federal agency said it would enforce the program , promising that consumers would soon see some reduction in telephone sales pitches .	
Ground Truth: fcc steps in to enforce do-not-call list ; bush signs new law to support program	(A-1)Supervised Result: second federal agency vows to reduction telephone sales
(C-2)WGAN: fight between government over national list took turn	(C-3)Adversarial REINFORCE: fight between government over national list took turn second
(E-2)WGAN : three stores and markets was forced to shut down	(E-3)Adversarial REINFORCE: fight between the government over the national list took another turn second

Figure 9

Source Text: the fbi has arrested a father and son with suspected links to al-qaeda in lodi , california , the los angeles times reported wednesday .	
Ground Truth: two men with suspected al-qaeda links arrested in california	(A-1)Supervised Result: fbi arrests father son with suspected links to al-qaeda
(C-2)WGAN: fbi arrested in father son with al-qaeda links	(C-3)Adversarial REINFORCE: fbi arrested father and son with suspected links to al-qaeda california
(E-2)WGAN : fbi has arrested father and son with suspected links to al-qaeda	(E-3)Adversarial REINFORCE: fbi has arrested a father and son with suspected links al-qaeda

Figure 7

Source Text: three stores and markets in beijing 's fengtai district have been forced to shut down and yesterday each was fined ###,### yuan -lrb- ##,### us dollars -rrb- for violating laws and regulations on fire prevention and control .	
Ground Truth: stores markets punished for lack of fire controls	(A-1)Supervised Result: three stores fined for violating on fire
(C-2)WGAN: three stores markets in beijing 's district have forced to shut down yesterday	(C-3)Adversarial REINFORCE: three stores markets in beijing 's district have forced to shut down
(E-2)WGAN : three stores and markets was forced to shut down and was fined	(E-3)Adversarial REINFORCE: three stores and markets in beijing 's district was fined

Figure 10

<p>Source text: lrb cnn rrb civil unions between people of the same sex will soon be recognized in chile . the country joined several of its south american neighbors in allowing the unions when president michelle bachelet enacted a new law on monday . this is a concrete step in the drive to end the difference between homosexual and heterosexual couples , bachelet said . the new law will take effect in six months . it will give legal weight to cohabiting relationships between two people of the same sex and between a man and a woman . the chilean government estimates that around 2 million people will be able to benefit from the change . the law is intended to end discrimination faced by commonlaw couples , such as not being allowed to visit partners in hospital , make medical decisions on their behalf or decide what to do with their remains . it also gives the couples greater rights in the realms of property , health care , pensions and inheritance . a number of south american nations have moved to allow samesex civil unions in recent years . but marriage between people of the same sex is legal only in argentina , brazil and uruguay .</p>
<p>Ground Truth: president michelle bachelet signs law that will take effect in six months . chile joins several other south american nations that allow the unions .</p>
<p>(C-2)WGAN: civil unions between people of the same sex will soon be recognized in chile . the country joined several of its south american neighbors in allowing the unions when president michelle enacted a new law on monday . this is a concrete step in the drive .</p>
<p>(C-3)Adversarial REINFORCE: president michelle enacted a new law on monday . this is a concrete step in the same sex will soon be recognized in chile . the country joined several of its south american neighbors .</p>

Figure 11: An example of generated summary of CNN/Diary Mail.