Unpaired Sentiment-to-Sentiment Translation: A Cycled Reinforcement Learning Approach

Jingjing Xu[1][*], **Xu Sun**[1][*], **Qi Zeng**[1], **Xuancheng Ren**[1],
Xiaodong Zhang[1], **Houfeng Wang**[1], **Wenjie Li**[2]

[1]MOE Key Lab of Computational Linguistics, School of EECS, Peking University
[2]Department of Computing, Hong Kong Polytechnic University
{jingjingxu,xusun,pkuzengqi,renxc,zxdcs,wanghf}@pku.edu.cn
cswjli@comp.polyu.edu.hk

Abstract

The goal of sentiment-to-sentiment "translation" is to change the underlying sentiment of a sentence while keeping its content. The main challenge is the lack of parallel data. To solve this problem, we propose a cycled reinforcement learning method that enables training on unpaired data by collaboration between a neutralization module and an emotionalization module. We evaluate our approach on two review datasets, Yelp and Amazon. Experimental results show that our approach significantly outperforms the state-of-the-art systems. Especially, the proposed method substantially improves the content preservation performance. The BLEU score is improved from 1.64 to 22.46 and from 0.56 to 14.06 on the two datasets, respectively.[1]

1 Introduction

Sentiment-to-sentiment "translation" requires the system to change the underlying sentiment of a sentence while preserving its non-emotional semantic content as much as possible. It can be regarded as a special style transfer task that is important in Natural Language Processing (NLP) (Hu et al., 2017; Shen et al., 2017; Fu et al., 2018). It has broad applications, including review sentiment transformation, news rewriting, etc. Yet the lack of parallel training data poses a great obstacle to a satisfactory performance.

Recently, several related studies for language style transfer (Hu et al., 2017; Shen et al., 2017) have been proposed. However, when applied to the sentiment-to-sentiment "translation" task, most existing studies only change the underlying sentiment and fail in keeping the semantic content. For example, given "The food is delicious" as the source input, the model generates "What a bad movie" as the output. Although the sentiment is successfully transformed from positive to negative, the output text focuses on a different topic. The reason is that these methods attempt to implicitly separate the emotional information from the semantic information in the same dense hidden vector, where all information is mixed together in an uninterpretable way. Due to the lack of supervised parallel data, it is hard to only modify the underlying sentiment without any loss of the non-emotional semantic information.

To tackle the problem of lacking parallel data, we propose a cycled reinforcement learning approach that contains two parts: a neutralization module and an emotionalization module. The neutralization module is responsible for extracting non-emotional semantic information by explicitly filtering out emotional words. The advantage is that only emotional words are removed, which does not affect the preservation of non-emotional words. The emotionalization module is responsible for adding sentiment to the neutralized semantic content for sentiment-to-sentiment translation.

In cycled training, given an emotional sentence with sentiment s, we first neutralize it to the non-emotional semantic content, and then force the emotionalization module to reconstruct the original sentence by adding the sentiment s. Therefore, the emotionalization module is taught to add sentiment to the semantic context in a supervised way. By adding opposite sentiment, we can achieve the goal of sentiment-to-sentiment translation. Because of the discrete choice of neutral words, the gradient is no longer differentiable over the neutralization module. Thus, we use policy gradient,

---

[*]Equal Contribution.
[1]The released code can be found in https://github.com/lancopku/unpaired-sentiment-translation

one of the reinforcement learning methods, to reward the output of the neutralization module based on the feedback from the emotionalization module. We add different sentiment to the semantic content and use the quality of the generated text as reward. The quality is evaluated by two useful metrics: one for identifying whether the generated text matches the target sentiment; one for evaluating the content preservation performance. The reward guides the neutralization module to better identify non-emotional words. In return, the improved neutralization module further enhances the emotionalization module.

Our contributions are concluded as follows:

- For sentiment-to-sentiment translation, we propose a cycled reinforcement learning approach. It enables training with unpaired data, in which only reviews and sentiment labels are available.

- Our approach tackles the bottleneck of keeping semantic information by explicitly separating sentiment information from semantic content.

- Experimental results show that our approach significantly outperforms the state-of-the-art systems, especially in content preservation.

2 Related Work

Style transfer in computer vision has been studied (Johnson et al., 2016; Gatys et al., 2016; Liao et al., 2017; Li et al., 2017; Zhu et al., 2017). The main idea is to learn the mapping between two image domains by capturing shared representations or correspondences of higher-level structures.

There have been some studies on unpaired language style transfer recently. Hu et al. (2017) propose a new neural generative model that combines variational auto-encoders (VAEs) and holistic attribute discriminators for effective imposition of style semantic structures. Fu et al. (2018) propose to use an adversarial network to make sure that the input content does not have style information. Shen et al. (2017) focus on separating the underlying content from style information. They learn an encoder that maps the original sentence to style-independent content and a style-dependent decoder for rendering. However, their evaluations only consider the transferred style accuracy.

We argue that content preservation is also an indispensable evaluation metric. However, when applied to the sentiment-to-sentiment translation task, the previously mentioned models share the same problem. They have the poor preservation of non-emotional semantic content.

In this paper, we propose a cycled reinforcement learning method to improve sentiment-to-sentiment translation in the absence of parallel data. The key idea is to build supervised training pairs by reconstructing the original sentence. A related study is "back reconstruction" in machine translation (He et al., 2016; Tu et al., 2017). They couple two inverse tasks: one is for translating a sentence in language A to a sentence in language B; the other is for translating a sentence in language B to a sentence in language A. Different from the previous work, we do not introduce the inverse task, but use collaboration between the neutralization module and the emotionalization module.

Sentiment analysis is also related to our work (Socher et al., 2011; Pontiki et al., 2015; Rosenthal et al., 2017; Chen et al., 2017; Ma et al., 2017, 2018b). The task usually involves detecting whether a piece of text expresses positive, negative, or neutral sentiment. The sentiment can be general or about a specific topic.

3 Cycled Reinforcement Learning for Unpaired Sentiment-to-Sentiment Translation

In this section, we introduce our proposed method. An overview is presented in Section 3.1. The details of the neutralization module and the emotionalization module are shown in Section 3.2 and Section 3.3. The cycled reinforcement learning mechanism is introduced in Section 3.4.

3.1 Overview

The proposed approach contains two modules: a neutralization module and an emotionalization module, as shown in Figure 1. The neutralization module first extracts non-emotional semantic content, and then the emotionalization module attaches sentiment to the semantic content. Two modules are trained by the proposed cycled reinforcement learning method. The proposed method requires the two modules to have initial learning ability. Therefore, we propose a novel pre-training method, which uses a self-attention based senti-
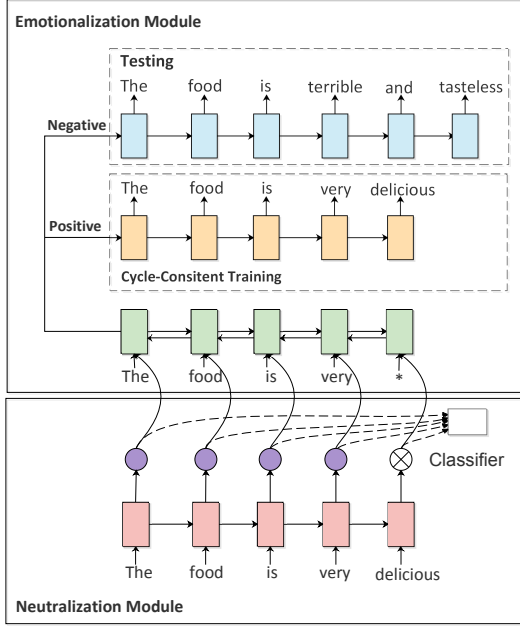
Figure 1: An illustration of the two modules. Lower: The neutralization module removes emotional words and extracts non-emotional semantic information. Upper: The emotionalization module adds sentiment to the semantic content. The proposed self-attention based sentiment classifier is used to guide the pre-training.

ment classifier (SASC). A sketch of cycled reinforcement learning is shown in Algorithm 1. The details are introduced as follows.

3.2 Neutralization Module

The neutralization module N_θ is used for explicitly filtering out emotional information. In this paper, we consider this process as an extraction problem. The neutralization module first identifies non-emotional words and then feeds them into the emotionalization module. We use a single Long-short Term Memory Network (LSTM) to generate the probability of being neutral or being polar for every word in a sentence. Given an emotional input sequence $x = (x_1, x_2, \ldots, x_T)$ of T words from Γ, the vocabulary of words, this module is responsible for producing a neutralized sequence.

Since cycled reinforcement learning requires the modules with initial learning ability, we propose a novel pre-training method to teach the neutralization module to identify non-emotional words. We construct a self-attention based sentiment classifier and use the learned attention weight as the supervisory signal. The motivation comes from the fact that, in a well-trained senti-

Algorithm 1 The cycled reinforcement learning method for training the neutralization module N_θ and the emotionalization module E_ϕ.

1: Initialize the neutralization module N_θ, the emotionalization module E_ϕ with random weights θ, ϕ
2: Pre-train N_θ using MLE based on Eq. 6
3: Pre-train E_ϕ using MLE based on Eq. 7
4: **for** each iteration $i = 1, 2, ..., M$ **do**
5:     Sample a sequence x with sentiment s from X
6:     Generate a neutralized sequence $\hat{x}$ based on N_θ
7:     Given $\hat{x}$ and s, generate an output based on E_ϕ
8:     Compute the gradient of E_ϕ based on Eq. 8
9:     Compute the reward R_1 based on Eq. 11
10:     $\bar{s}$ = the opposite sentiment
11:     Given $\hat{x}$ and $\bar{s}$, generate an output based on E_ϕ
12:     Compute the reward R_2 based on Eq. 11
13:     Compute the combined reward R_c based on Eq. 10
14:     Compute the gradient of N_θ based on Eq. 9
15:     Update model parameters θ, ϕ
16: **end for**

---

ment classification model, the attention weight reflects the sentiment contribution of each word to some extent. Emotional words tend to get higher attention weights while neutral words usually get lower weights. The details of sentiment classifier are described as follows.

Given an input sequence x, a sentiment label y is produced as

$$y = softmax(W \cdot \boldsymbol{c}) \qquad (1)$$

where W is a parameter. The term $\boldsymbol{c}$ is computed as a weighted sum of hidden vectors:

$$\boldsymbol{c} = \sum_{i=0}^{T} \alpha_i \boldsymbol{h}_i \qquad (2)$$

where α_i is the weight of h_i. The term $\boldsymbol{h}_i$ is the output of LSTM at the i-th word. The term α_i is computed as

$$\alpha_i = \frac{\exp(e_i)}{\sum_{i=0}^{T} \exp(e_i)} \qquad (3)$$

where $e_i = f(\boldsymbol{h}_i, \boldsymbol{h}_T)$ is an alignment model. We consider the last hidden state $\boldsymbol{h}_T$ as the context vector, which contains all information of an input sequence. The term e_i evaluates the contribution of each word for sentiment classification.

Our experimental results show that the proposed sentiment classifier achieves the accuracy of 89% and 90% on two datasets. With high classification accuracy, the attention weight produced by the classifier is considered to adequately capture the sentiment information of each word.

To extract non-emotional words based on continuous attention weights, we map attention weights to discrete values, 0 and 1. Since the discrete method is not the key part is this paper, we only use the following method for simplification.

We first calculate the averaged attention value in a sentence as

$$\bar{\alpha} = \frac{1}{T} \sum_{i=0}^{T} \alpha_i \tag{4}$$

where $\bar{\alpha}$ is used as the threshold to distinguish non-emotional words from emotional words. The discrete attention weight is calculated as

$$\hat{\alpha}_i = \begin{cases} 1, & \text{if } \alpha_i \leq \bar{\alpha} \\ 0, & \text{if } \alpha_i > \bar{\alpha} \end{cases} \tag{5}$$

where $\hat{\alpha}_i$ is treated as the identifier.

For pre-training the neutralization module, we build the training pair of input text x and a discrete attention weight sequence $\hat{\alpha}$. The cross entropy loss is computed as

$$L_\theta = - \sum_{i=1}^{T} P_{N_\theta}(\hat{\alpha}_i | x_i) \tag{6}$$

3.3 Emotionalization Module

The emotionalization module E_ϕ is responsible for adding sentiment to the neutralized semantic content. In our work, we use a bi-decoder based encoder-decoder framework, which contains one encoder and two decoders. One decoder adds the positive sentiment and the other adds the negative sentiment. The input sentiment signal determines which decoder to use. Specifically, we use the seq2seq model (Sutskever et al., 2014) for implementation. Both the encoder and decoder are LSTM networks. The encoder learns to compress the semantic content into a dense vector. The decoder learns to add sentiment based on the dense vector. Given the neutralized semantic content and the target sentiment, this module is responsible for producing an emotional sequence.

For pre-training the emotionalization module, we first generate a neutralized input sequence $\hat{x}$ by removing emotional words identified by the proposed sentiment classifier. Given the training pair of a neutralized sequence $\hat{x}$ and an original sentence x with sentiment s, the cross entropy loss is computed as

$$L_\phi = - \sum_{i=1}^{T} P_{E_\phi}(x_i | \hat{x}_i, s) \tag{7}$$

where a positive example goes through the positive decoder and a negative example goes through the negative decoder.

We also explore a simpler method for pre-training the emotionalization module, which uses the product between a continuous vector $1 - \alpha$ and a word embedding sequence as the neutralized content where α represents an attention weight sequence. Experimental results show that this method achieves much lower results than explicitly removing emotional words based on discrete attention weights. Thus, we do not choose this method in our work.

3.4 Cycled Reinforcement Learning

Two modules are trained by the proposed cycled method. The neutralization module first neutralizes an emotional input to semantic content and then the emotionalization module is forced to reconstruct the original sentence based on the source sentiment and the semantic content. Therefore, the emotionalization module is taught to add sentiment to the semantic content in a supervised way. Because of the discrete choice of neutral words, the loss is no longer differentiable over the neutralization module. Therefore, we formulate it as a reinforcement learning problem and use policy gradient to train the neutralization module. The detailed training process is shown as follows.

We refer the neutralization module N_θ as the first agent and the emotionalization module E_ϕ as the second one. Given a sentence x associated with sentiment s, the term $\hat{x}$ represents the middle neutralized context extracted by $\hat{\alpha}$, which is generated by $P_{N_\theta}(\hat{\alpha}|x)$.

In cycled training, the original sentence can be viewed as the supervision for training the second agent. Thus, the gradient for the second agent is

$$\nabla_\phi J(\phi) = \nabla_\phi \log(P_{E_\phi}(x|\hat{x}, s)) \tag{8}$$

We denote $\bar{x}$ as the output generated by $P_{E_\phi}(\bar{x}|\hat{x}, s)$. We also denote y as the output generated by $P_{E_\phi}(y|\hat{x}, \bar{s})$ where $\bar{s}$ represents the opposite sentiment. Given $\bar{x}$ and y, we first calculate rewards for training the neutralized module, R_1 and R_2. The details of calculation process will be introduced in Section 3.4.1. Then, we optimize parameters through policy gradient by maximizing the expected reward to train the neutralization module. It guides the neutralization module to identify non-emotional words better. In return, the

improved neutralization module further enhances the emotionalization module.

According to the policy gradient theorem (Williams, 1992), the gradient for the first agent is

$$\nabla_\theta J(\theta) = \mathbb{E}[R_c \cdot \nabla_\theta \log(P_{N_\theta}(\hat{\boldsymbol{\alpha}}|\boldsymbol{x}))] \quad (9)$$

where R_c is calculated as

$$R_c = R_1 + R_2 \quad (10)$$

Based on Eq. 8 and Eq. 9, we use the sampling approach to estimate the expected reward. This cycled process is repeated until converge.

3.4.1 Reward

The reward consists of two parts, sentiment confidence and BLEU. Sentiment confidence evaluates whether the generated text matches the target sentiment. We use a pre-trained classifier to make the judgment. Specially, we use the proposed self-attention based sentiment classifier for implementation. The BLEU (Papineni et al., 2002) score is used to measure the content preservation performance. Considering that the reward should encourage the model to improve both metrics, we use the harmonic mean of sentiment confidence and BLEU as reward, which is formulated as

$$R = (1 + \beta^2)\frac{2 \cdot BLEU \cdot Confid}{(\beta^2 \cdot BLEU) + Confid} \quad (11)$$

where β is a harmonic weight.

4 Experiment

In this section, we evaluate our method on two review datasets. We first introduce the datasets, the training details, the baselines, and the evaluation metrics. Then, we compare our approach with the state-of-the-art systems. Finally, we show the experimental results and provide the detailed analysis of the key components.

4.1 Unpaired Datasets

We conduct experiments on two review datasets that contain user ratings associated with each review. Following previous work (Shen et al., 2017), we consider reviews with rating above three as positive reviews and reviews below three as negative reviews. The positive and negative reviews are not paired. Since our approach focuses on sentence-level sentiment-to-sentiment translation where sentiment annotations are provided at the document level, we process the two datasets with the following steps. First, following previous work (Shen et al., 2017), we filter out the reviews that exceed 20 words. Second, we construct text-sentiment pairs by extracting the first sentence in a review associated with its sentiment label, because the first sentence usually expresses the core idea. Finally, we train a sentiment classifier and filter out the text-sentiment pairs with the classifier confidence below 0.8. Specially, we use the proposed self-attention based sentiment classifier for implementation. The details of the processed datasets are introduced as follows.

Yelp Review Dataset (Yelp): This dataset is provided by Yelp Dataset Challenge.[2] The processed Yelp dataset contains 400K, 10K, and 3K pairs for training, validation, and testing, respectively.

Amazon Food Review Dataset (Amazon): This dataset is provided by McAuley and Leskovec (2013). It consists of amounts of food reviews from Amazon.[3] The processed Amazon dataset contains 230K, 10K, and 3K pairs for training, validation, and testing, respectively.

4.2 Training Details

We tune hyper-parameters based on the performance on the validation sets. The self-attention based sentiment classifier is trained for 10 epochs on two datasets. We set β for calculating reward to 0.5, hidden size to 256, embedding size to 128, vocabulary size to 50K, learning rate to 0.6, and batch size to 64. We use the Adagrad (Duchi et al., 2011) optimizer. All of the gradients are clipped when the norm exceeds 2. Before cycled training, the neutralization module and the emotionalization module are pre-trained for 1 and 4 epochs on the yelp dataset, for 3 and 5 epochs on the Amazon dataset.

4.3 Baselines

We compare our proposed method with the following state-of-the-art systems.

Cross-Alignment Auto-Encoder (CAAE): This method is proposed by Shen et al. (2017). They propose a method that uses refined alignment of latent representations in hidden layers to

---

[2]https://www.yelp.com/dataset/challenge
[3]http://amazon.com

perform style transfer. We treat this model as a baseline and adapt it by using the released code.

Multi-Decoder with Adversarial Learning (MDAL): This method is proposed by Fu et al. (2018). They use a multi-decoder model with adversarial learning to separate style representations and content representations in hidden layers. We adapt this model by using the released code.

4.4 Evaluation Metrics

We conduct two evaluations in this work, including an automatic evaluation and a human evaluation. The details of evaluation metrics are shown as follows.

4.4.1 Automatic Evaluation

We quantitatively measure sentiment transformation by evaluating the accuracy of generating designated sentiment. For a fair comparison, we do not use the proposed sentiment classification model. Following previous work (Shen et al., 2017; Hu et al., 2017), we instead use a state-of-the-art sentiment classifier (Vieira and Moura, 2017), called TextCNN, to automatically evaluate the transferred sentiment accuracy. TextCNN achieves the accuracy of 89% and 88% on two datasets. Specifically, we generate sentences given sentiment s, and use the pre-trained sentiment classifier to assign sentiment labels to the generated sentences. The accuracy is calculated as the percentage of the predictions that match the sentiment s.

To evaluate the content preservation performance, we use the BLEU score (Papineni et al., 2002) between the transferred sentence and the source sentence as an evaluation metric. BLEU is a widely used metric for text generation tasks, such as machine translation, summarization, etc. The metric compares the automatically produced text with the reference text by computing overlapping lexical n-gram units.

To evaluate the overall performance, we use the geometric mean of ACC and BLEU as an evaluation metric. The G-score is one of the most commonly used "single number" measures in Information Retrieval, Natural Language Processing, and Machine Learning.

4.4.2 Human Evaluation

While the quantitative evaluation provides indication of sentiment transfer quality, it can not evaluate the quality of transferred text accurately.

Yelp	ACC	BLEU	G-score
CAAE (Shen et al., 2017)	93.22	1.17	10.44
MDAL (Fu et al., 2018)	85.65	1.64	11.85
Proposed Method	80.00	22.46	**42.38**
Amazon	ACC	BLEU	G-score
CAAE (Shen et al., 2017)	84.19	0.56	6.87
MDAL (Fu et al., 2018)	70.50	0.27	4.36
Proposed Method	70.37	14.06	**31.45**

Table 1: Automatic evaluations of the proposed method and baselines. ACC evaluates sentiment transformation. BLEU evaluates content preservation. G-score is the geometric mean of ACC and BLEU.

Therefore, we also perform a human evaluation on the test set. We randomly choose 200 items for the human evaluation. Each item contains the transformed sentences generated by different systems given the same source sentence. The items are distributed to annotators who have no knowledge about which system the sentence is from. They are asked to score the transformed text in terms of sentiment and semantic similarity. Sentiment represents whether the sentiment of the source text is transferred correctly. Semantic similarity evaluates the context preservation performance. The score ranges from 1 to 10 (1 is very bad and 10 is very good).

4.5 Experimental Results

Automatic evaluation results are shown in Table 1. ACC evaluates sentiment transformation. BLEU evaluates semantic content preservation. G-score represents the geometric mean of ACC and BLEU. CAAE and MDAL achieve much lower BLEU scores, 1.17 and 1.64 on the Yelp dataset, 0.56 and 0.27 on the Amazon dataset. The low BLEU scores indicate the worrying content preservation performance to some extent. Even with the desired sentiment, the irrelevant generated text leads to worse overall performance. In general, these two systems work more like sentiment-aware language models that generate text only based on the target sentiment and neglect the source input. The main reason is that these two systems attempt to separate emotional information from non-emotional content in a hidden layer, where all information is complicatedly mixed together. It is difficult to only modify emotional information without any loss of non-emotional semantic content.

In comparison, our proposed method achieves the best overall performance on the two datasets,

Yelp	Sentiment	Semantic	G-score
CAAE (Shen et al., 2017)	7.67	3.87	5.45
MDAL (Fu et al., 2018)	7.12	3.68	5.12
Proposed Method	6.99	5.08	**5.96**

Amazon	Sentiment	Semantic	G-score
CAAE (Shen et al., 2017)	8.61	3.15	5.21
MDAL (Fu et al., 2018)	7.93	3.22	5.05
Proposed Method	7.92	4.67	**6.08**

Table 2: Human evaluations of the proposed method and baselines. *Sentiment* evaluates sentiment transformation. *Semantic* evaluates content preservation.

demonstrating the ability of learning knowledge from unpaired data. This result is attributed to the improved BLEU score. The BLEU score is largely improved from 1.64 to 22.46 and from 0.56 to 14.06 on the two datasets. The score improvements mainly come from the fact that we separate emotional information from semantic content by explicitly filtering out emotional words. The extracted content is preserved and fed into the emotionalization module. Given the overall quality of transferred text as the reward, the neutralization module is taught to extract non-emotional semantic content better.

Table 2 shows the human evaluation results. It can be clearly seen that the proposed method obviously improves semantic preservation. The semantic score is increased from 3.87 to 5.08 on the Yelp dataset, and from 3.22 to 4.67 on the Amazon dataset. In general, our proposed model achieves the best overall performance. Furthermore, it also needs to be noticed that with the large improvement in content preservation, the sentiment accuracy of the proposed method is lower than that of CAAE on the two datasets. It shows that simultaneously promoting sentiment transformation and content preservation remains to be studied further.

By comparing two evaluation results, we find that there is an agreement between the human evaluation and the automatic evaluation. It indicates the usefulness of automatic evaluation metrics. However, we also notice that the human evaluation has a smaller performance gap between the baselines and the proposed method than the automatic evaluation. It shows the limitation of automatic metrics for giving accurate results. For evaluating sentiment transformation, even with a high accuracy, the sentiment classifier sometimes generates noisy results, especially for those neutral sentences (e.g., "I ate a cheese sandwich"). For evaluating content preservation, the BLEU score

Input: *I would strongly advise against using this company.*
CAAE: *I love this place for a great experience here.*
MDAL: *I have been a great place was great.*
Proposed Method: *I would love using this company.*

Input: *The service was nearly non-existent and extremely rude.*
CAAE: *The best place in the best area in vegas.*
MDAL: *The food is very friendly and very good.*
Proposed Method: *The service was served and completely fresh.*

Input: *Asked for the roast beef and mushroom sub, only received roast beef.*
CAAE: *We had a great experience with.*
MDAL: *This place for a great place for a great food and best.*
Proposed Method: *Thanks for the beef and spring bbq.*

Input: *Worst cleaning job ever!*
CAAE: *Great food and great service!*
MDAL: *Great food, food!*
Proposed Method: *Excellent outstanding job ever!*

Input: *Most boring show I've ever been.*
CAAE: *Great place is the best place in town.*
MDAL: *Great place I've ever ever had.*
Proposed Method: *Most amazing show I've ever been.*

Input: *Place is very clean and the food is delicious.*
CAAE: *Don't go to this place.*
MDAL: *This place wasn't worth the worst place is horrible.*
Proposed Method: *Place is very small and the food is terrible.*

Input: *Really satisfied with experience buying clothes.*
CAAE: *Don't go to this place.*
MDAL: *Do not impressed with this place.*
Proposed Method: *Really bad experience.*

Table 3: Examples generated by the proposed approach and baselines on the Yelp dataset. The two baselines change not only the polarity of examples, but also the semantic content. In comparison, our approach changes the sentiment of sentences with higher semantic similarity.

is computed based on the percentage of overlapping n-grams between the generated text and the reference text. However, the overlapping n-grams contain not only content words but also function words, bringing the noisy results. In general, accurate automatic evaluation metrics are expected in future work.

Table 3 presents the examples generated by different systems on the Yelp dataset. The two baselines change not only the polarity of examples, but also the semantic content. In comparison, our method precisely changes the sentiment of sentences (and paraphrases slightly to ensure fluency), while keeping the semantic content unchanged.

Yelp	ACC	BLEU	G-score
Emotionalization Module	41.84	25.66	32.77
+ NM + Cycled RL	85.71	1.08	9.62
+ NM + Pre-training	70.61	17.02	34.66
+ NM + Cycled RL + Pre-training	80.00	22.46	**42.38**
Amazon	ACC	BLEU	G-score
Emotionalization Module	57.28	12.22	26.46
+ NM + Cycled RL	64.16	8.03	22.69
+ NM + Pre-training	69.61	11.16	27.87
+ NM + Cycled RL + Pre-training	70.37	14.06	**31.45**

Table 4: Performance of key components in the proposed approach. "NM" denotes the neutralization module. "Cycled RL" represents cycled reinforcement learning.

4.6 Incremental Analysis

In this section, we conduct a series of experiments to evaluate the contributions of our key components. The results are shown in Table 4.

We treat the emotionalization module as a baseline where the input is the original emotional sentence. The emotionalization module achieves the highest BLEU score but with much lower sentiment transformation accuracy. The encoding of the original sentiment leads to the emotional hidden vector that influences the decoding process and results in worse sentiment transformation performance.

It can be seen that the method with all components achieves the best performance. First, we find that the method that only uses cycled reinforcement learning performs badly because it is hard to guide two randomly initialized modules to teach each other. Second, the pre-training method brings a slight improvement in overall performance. The G-score is improved from 32.77 to 34.66 and from 26.46 to 27.87 on the two datasets. The bottleneck of this method is the noisy attention weight because of the limited sentiment classification accuracy. Third, the method that combines cycled reinforcement learning and pre-training achieves the better performance than using one of them. Pre-training gives the two modules initial learning ability. Cycled training teaches the two modules to improve each other based on the feedback signals. Specially, the G-score is improved from 34.66 to 42.38 and from 27.87 to 31.45 on the two datasets. Finally, by comparing the methods with and without the neutralization module, we find that the neutralization mechanism improves a lot in sentiment transformation with a slight reduction on content preservation. It proves the effectiveness of explic-

Michael is absolutely *wonderful.*
I would strongly advise against *using this company.*
Horrible experience!
Worst cleaning job ever!
Most boring *show i 've ever been.*
Hainan chicken was really good.
I really don't understand all the negative reviews *for this dentist.*
Smells so weird *in there.*
The service was nearly non-existent *and extremely* rude.

Table 5: Analysis of the neutralization module. Words in red are removed by the neutralization module.

itly separating sentiment information from semantic content.

Furthermore, to analyze the neutralization ability in the proposed method, we randomly sample several examples, as shown in Table 5. It can be clearly seen that emotional words are removed accurately almost without loss of non-emotional information.

4.7 Error Analysis

Although the proposed method outperforms the state-of-the-art systems, we also observe several failure cases, such as sentiment-conflicted sentences (e.g., "Outstanding and bad service"), neutral sentences (e.g., "Our first time here"). Sentiment-conflicted sentences indicate that the original sentiment is not removed completely. This problem occurs when the input contains emotional words that are unseen in the training data, or the sentiment is implicitly expressed. Handling complex sentiment expressions is an important problem for future work. Neutral sentences demonstrate that the decoder sometimes fails in adding the target sentiment and only generates text based on the semantic content. A better sentiment-aware decoder is expected to be explored in future work.

5 Conclusions and Future Work

In this paper, we focus on unpaired sentiment-to-sentiment translation and propose a cycled reinforcement learning approach that enables training in the absence of parallel training data. We conduct experiments on two review datasets. Experimental results show that our method substantially outperforms the state-of-the-art systems, especially in terms of semantic preservation. For future work, we would like to explore a fine-grained version of sentiment-to-sentiment translation that

not only reverses sentiment, but also changes the strength of sentiment.

References

Tao Chen, Ruifeng Xu, Yulan He, and Xuan Wang. 2017. Improving sentiment analysis via sentence type classification using bilstm-crf and CNN. *Expert Syst. Appl.*, 72:221–230.

Li Dong, Shaohan Huang, Furu Wei, Mirella Lapata, Ming Zhou, and Ke Xu. 2017. Learning to generate product reviews from attributes. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, volume 1, pages 623–632.

John C. Duchi, Elad Hazan, and Yoram Singer. 2011. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12:2121–2159.

Zhenxin Fu, Xiaoye Tan, Nanyun Peng, Dongyan Zhao, and Rui Yan. 2018. Style transfer in text: Exploration and evaluation. In *AAAI 2018*.

Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. 2016. Image style transfer using convolutional neural networks. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 2414–2423.

Di He, Yingce Xia, Tao Qin, Liwei Wang, Nenghai Yu, Tie-Yan Liu, and Wei-Ying Ma. 2016. Dual learning for machine translation. In *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, pages 820–828.

Zhiting Hu, Zichao Yang, Xiaodan Liang, Ruslan Salakhutdinov, and Eric P. Xing. 2017. Controllable text generation. In *ICML 2017*.

Justin Johnson, Alexandre Alahi, and Li Fei-Fei. 2016. Perceptual losses for real-time style transfer and super-resolution. In *ECCV, 2016*, pages 694–711.

Yanghao Li, Naiyan Wang, Jiaying Liu, and Xiaodi Hou. 2017. Demystifying neural style transfer. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*, pages 2230–2236.

Jing Liao, Yuan Yao, Lu Yuan, Gang Hua, and Sing Bing Kang. 2017. Visual attribute transfer through deep image analogy. *ACM Trans. Graph.*, 36(4):120:1–120:15.

Junyang Lin, Shuming Ma, Qi Su, and Xu Sun. 2018. Decoding-history-based adaptive control of attention for neural machine translation. *CoRR*, abs/1802.01812.

Dehong Ma, Sujian Li, Xiaodong Zhang, Houfeng Wang, and Xu Sun. 2017. Cascading multiway attentions for document-level sentiment classification. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing, IJCNLP 2017, Taipei, Taiwan, November 27 - December 1, 2017 - Volume 1: Long Papers*, pages 634–643.

Shuming Ma, Xu Sun, Wei Li, Sujian Li, Wenjie Li, and Xuancheng Ren. 2018a. Query and output: Generating words by querying distributed word representations for paraphrase generation. *CoRR*, abs/1803.01465.

Shuming Ma, Xu Sun, Junyang Lin, and Xuancheng Ren. 2018b. A hierarchical end-to-end model for jointly improving text summarization and sentiment classification. *CoRR*, abs/1805.01089.

Julian John McAuley and Jure Leskovec. 2013. From amateurs to connoisseurs: modeling the evolution of user expertise through online reviews. In *22nd International World Wide Web Conference, WWW '13, Rio de Janeiro, Brazil, May 13-17, 2013*, pages 897–908.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, July 6-12, 2002, Philadelphia, PA, USA.*, pages 311–318.

Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Suresh Manandhar, and Ion Androutsopoulos. 2015. Semeval-2015 task 12: Aspect based sentiment analysis. In *Proceedings of the 9th International Workshop on Semantic Evaluation, SemEval@NAACL-HLT 2015, Denver, Colorado, USA, June 4-5, 2015*, pages 486–495.

Sara Rosenthal, Noura Farra, and Preslav Nakov. 2017. Semeval-2017 task 4: Sentiment analysis in twitter. In *Proceedings of the 11th International Workshop on Semantic Evaluation, SemEval@ACL 2017, Vancouver, Canada, August 3-4, 2017*, pages 502–518.

Tianxiao Shen, Tao Lei, Regina Barzilay, and Tommi S. Jaakkola. 2017. Style transfer from non-parallel text by cross-alignment. In *NIPS 2017*.

Richard Socher, Cliff Chiung-Yu Lin, Andrew Y. Ng, and Christopher D. Manning. 2011. Parsing natural scenes and natural language with recursive neural networks. In *Proceedings of the 28th International Conference on Machine Learning, ICML 2011, Bellevue, Washington, USA, June 28 - July 2, 2011*, pages 129–136.

Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. 2014. Sequence to sequence learning with neural networks. In *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, pages 3104–3112.

Zhaopeng Tu, Yang Liu, Lifeng Shang, Xiaohua Liu, and Hang Li. 2017. Neural machine translation with reconstruction. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA.*, pages 3097–3103.

Joao Paulo Albuquerque Vieira and Raimundo Santos Moura. 2017. An analysis of convolutional neural networks for sentence classification. In *XLIII 2017*, pages 1–5.

Ronald J. Williams. 1992. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 8:229–256.

Jingjing Xu, Xu Sun, Xuancheng Ren, Junyang Lin, Bingzhen Wei, and Wei Li. 2018. DP-GAN: diversity-promoting generative adversarial network for generating informative and diversified text. *CoRR*, abs/1802.01345.

Hongyu Zang and Xiaojun Wan. 2017. Towards automatic generation of product reviews from aspect-sentiment scores. In *Proceedings of the 10th International Conference on Natural Language Generation*, pages 168–177.

Zhiyuan Zhang, Wei Li, and Xu Sun. 2018. Automatic transferring between ancient chinese and contemporary chinese. *CoRR*, abs/1803.01557.

Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*, pages 2242–2251.