

Collaborative Learning of Bidirectional Decoders for Unsupervised Text Style Transfer

Yun Ma¹, Yangbin Chen², Xudong Mao³, Qing Li¹

¹Department of Computing, The Hong Kong Polytechnic University

²Department of Information Engineering, The Chinese University of Hong Kong

³Department of Artificial Intelligence, Xiamen University

mayun371@gmail.com, dongyiwu92@gmail.com

xudong.xdmao@gmail.com, csqli@comp.polyu.edu.hk

Abstract

Unsupervised text style transfer aims to alter the underlying style of the text to a desired value while keeping its style-independent semantics, without the support of parallel training corpora. Existing methods struggle to achieve both high style conversion rate and low content loss, exhibiting the over-transfer and under-transfer problems. We attribute these problems to the conflicting driving forces of the style conversion goal and content preservation goal. In this paper, we propose a collaborative learning framework for unsupervised text style transfer using a pair of bidirectional decoders, one decoding from left to right while the other decoding from right to left. In our collaborative learning mechanism, each decoder is regularized by knowledge from its peer which has a different knowledge acquisition process. The difference is guaranteed by their opposite decoding directions and a distinguishability constraint. As a result, mutual knowledge distillation drives both decoders to a better optimum and alleviates the over-transfer and under-transfer problems. Experimental results on two benchmark datasets show that our framework achieves strong empirical results on both style compatibility and content preservation.

1 Introduction

Text style transfer is to transform an input text of a source style to a target style (i.e., *style conversion goal*) without loss of its style-independent information (i.e., *content preservation goal*). Concentrating on different stylistic attributes, text style transfer has attracted much attention from various natural language processing applications, such as personalized machine translation (Rabinovich et al., 2017), text formalization (Zhang et al., 2020), and sentiment translation (Xu et al., 2018). Unfortunately, the parallel corpora with aligned input and output are usually unavailable, challenging the models to learn in an unsupervised manner.

Input	The dish is fresh and yummy.
Expected Output	The dish is old and disgusting.
Over-Transfer	The staff are rude!
Under-Transfer	The dish is old and yummy.

Table 1: The over-transferred and the under-transferred results for an exemplar input in the positive→negative sentiment transfer task.

One research line to address the unsupervised text style transfer task is to first disentangle the style-independent semantics (content) from the style-dependent semantics (style), and then produce the output based on the disentangled content and the target style. The disentanglement is enforced either *implicitly* (Hu et al., 2017; Shen et al., 2017; Fu et al., 2018; John et al., 2019), or *explicitly* (Li et al., 2018; Wu et al., 2019b; Xu et al., 2018; Madaan et al., 2020). Nevertheless, such disentanglement has been discovered to be hardly met in practice (Elazar and Goldberg, 2018). Putting aside the disentanglement step, another research line learns a direct mapping from input to output, where the model is optimized by pseudo-parallel data created by online back-translation (Lample et al., 2019; Zhang et al., 2018c; Luo et al., 2019; Pant et al., 2020), or jointly predicting the word-level style relevance (Zhou et al., 2020). For both the disentanglement and non-disentanglement based research lines, objectives like self-reconstruction and style classification have been extensively proven as effective in guiding the training process.

Despite the great progress, existing methods still struggle to achieve both high style conversion rate and low content loss. Such limitation is widely embodied by the over-transfer and under-transfer problems: **Over-Transfer (OT)** refers to the content deviation patterns that some style-independent semantics are altered; while **Under-Transfer (UT)** refers to the lazy copying patterns that some style-dependent semantics are unchanged. Table 1 illustrates the OT and UT problems in a sentiment trans-

fer scenario. For the given input, the OT output achieves the positive-to-negative sentiment conversion but undesirably changes the focused aspect from *dish* to *staff*; in contrast, the UT output preserves this sentiment-independent content but fails to convert *yummy* to words with negative sentiment.

The OT and UT problems are the product of the conflicting driving forces of the style conversion goal and content preservation goal. Specifically, objectives for the style conversion goal (e.g., the style classification loss) encourage generating new words reflecting the target style; while objectives for the content preservation goal (e.g., the self-reconstruction loss) encourage copying from source words. Without supervision from ground-truth, the model struggles between these two conflicting forces and tends to put their probability mass on choices in both directions to achieve both goals. As a result, the model can make unconfident predictions and present the OT and UT problems when biasing in the wrong direction. Furthermore, the specific design of different methods may further exacerbate the OT / UT problems¹.

In this paper, we draw inspiration from multi-agent learning to address the OT and UT problems. Under the widely adopted encoder-decoder architecture, we jointly learn a pair of Collaborative Bidirectional Decoders (CBD), one decoding from left to right (L2R) and the other decoding from right to left (R2L). Our collaborative learning mechanism regularizes each decoder by distilling knowledge from its peer. Essentially, OT and UT problems are incorrectly predicted words in the decoding procedure. In a similar spirit of pseudo-labeling (Lee, 2013) and consistency regularization (Laine and Aila, 2017) in Semi-Supervised Learning (SSL), the mutual knowledge distillation provides a direct optimization direction for data lacking ground-truth, gradually improving both decoders to reduce OT and UT errors and get more peaked on reasonable predictions. Specifically, consistent predictions will be reinforced, while the inconsistent predictions lead to more uncertainty over candidate predictions and provide a chance for achieving consistency in subsequent training. However, this is only plausible under the consis-

tency assumption that consistent knowledge can represent the ground-truth with a high probability. As with the Co-Training framework (Blum and Mitchell, 1998; Qiao et al., 2018) in SSL, to guarantee the rationality of the consistency assumption, we require the two decoders to have different knowledge acquisition processes. In addition to their opposite decoding directions, we introduce a distinguishability constraint to ensure their difference. In particular, an additional discriminator is employed to distinguish the softmax probabilities from the two decoders.

Our contributions can be summarized as: (1) We address the over-transfer and under-transfer problems in unsupervised text style transfer from the perspective of multi-agent learning with a pair of bidirectional decoders. (2) We propose a collaborative learning mechanism with mutual knowledge distillation and a distinguishability constraint to optimize the bidirectional decoders, so as to continuously promote the model’s capability. (3) Experimental results and in-depth analysis on two benchmark datasets verify the strength of our model on persuing both the style conversion goal and the content preservation goal.

2 Related Work

2.1 Unsupervised Text Style Transfer

One main branch of approaches tries to learn a style-agnostic representation by disentangling the content and style. Fu et al. (2018) and John et al. (2019) align the latent content space with adversarial objectives to eliminate the style information. Shen et al. (2017) directly aligns the transferred samples with the real samples from the target style with adversarial discriminators. Yang et al. (2018) achieves the alignment with language models instead of the binary discriminators to provide token-level supervision. Prabhumoye et al. (2018) assumes the style-agnostic representation can be achieved by translating the input text to another language. With all the content reduced to a latent vector, such implicit disentanglement based methods are prone to semantic loss. To address this limitation, Li et al. (2018), Xu et al. (2018), and Zhang et al. (2018b) remove the style indicators based on term frequencies or attention scores, and treat the remaining words as the disentangled content for stylization. Focusing on politeness transfer, Madaan et al. (2020) generalizes this explicit disentanglement step by allowing inserting placeholders

¹The disentanglement based methods can produce imperfect disentangled representations with content information eliminated or with style information kept, corresponding to OT or PT problems. The back-translation in non-disentanglement based methods can strengthen the UT patterns occurring in pseudo data by iteratively feeding them to the model.

without removing other content, where the placeholders are further replaced with stylistically relevant words.

Alternatively, recent methods skip the disentanglement step and directly optimize the mapping function from input to output. Motivated by the unsupervised machine translation approaches, [Zhang et al. \(2018c\)](#) and [Lample et al. \(2019\)](#) train the model with pseudo-parallel data dynamically created by back-translation. [Luo et al. \(2019\)](#) further proposes to use the reinforcement learning technique with the style conversion reward and the content preservation reward. [Zhou et al. \(2020\)](#) binds the output generation with word-level style relevance prediction. Standing outside the sequence-to-sequence models, [Wu et al. \(2019a\)](#) considers style transfer as a text editing task and learns where and how to operate the input via hierarchical reinforcement learning. Meanwhile, some works ([Sudhakar et al., 2019](#); [Wu et al., 2019b](#); [Dai et al., 2019](#)) advance the field by replacing the Recurrent Neural Networks (RNN) based architectures with the Transformer architecture for its superiority in capturing long-term dependencies.

2.2 Multi-Agent Learning

Multi-agent learning improves model performance by incorporating multiple interactive agents. Most related to our work is the bidirectional decoding models ([Zhang et al., 2018a, 2019](#); [Zhou et al., 2019](#)) in Neural Machine Translation (NMT), which jointly train an L2R translator and an R2L one. [Zhang et al. \(2019\)](#) minimizes the KL divergence between the two translators to fuse the good prefixes of L2R decoding and the good suffixes of R2L decoding. [Bi et al. \(2019\)](#) further explores more than two agents where each agent learns the knowledge from a dynamic ensemble model. Mutual learning has also been set up between an NMT agent and a Statistical Machine Translation (SMT) agent to integrate NMT’s fluency and SMT’s robustness to noisy data ([Ren et al., 2019](#)). In addition to focusing on different tasks, our model differentiates by including a distinguishability constraint. Unlike NMT, the task of unsupervised text style transfer is not well-constrained due to both the lack of ground-truth and the conflicting forces from the two goals of style transfer. The distinguishability constraint is important to prevent the decoders from collapsing to one bad local optimum and reinforcing incorrect but consistent patterns.

3 Our Approach

Consider a training corpus $\mathcal{M} = \{(x_i, s_i)\}_{i=1}^N$, where x_i is a text sequence, and $s_i \in \mathcal{S}$ is its style with \mathcal{S} denoting all possible style types. The objective of text style transfer is to learn a conditional probability distribution $P(\tilde{x}|x, \tilde{s})$ to transform a given x to \tilde{x} with a target style \tilde{s} . The output \tilde{x} is expected to retain the style-independent information in x . Here, we stick to the encoder-decoder based sequence-to-sequence architecture, where an encoder E first encodes x to latent vectors $E(x)$, and a decoder D then produces \tilde{x} by sampling from its parameterized distribution $D(\tilde{x}|E(x), \tilde{s})$.

In this paper, we propose a framework with one encoder E and a pair of bidirectional decoders: an L2R decoder D^l producing the output sequence from left to right, and an R2L decoder D^r going in the opposite direction. The two decoders interact with each other in a collaborative learning mechanism. This collaborative learning mechanism is integrated with a basic framework that follows the non-disentanglement based research line and has three widely used objectives, i.e., self-reconstruction, back-translation, and style classification. In the following, we first briefly introduce the basic framework extended for our two-decoder scenario (Section 3.1), then elaborate our collaborative learning mechanism (Section 3.2), and present the training algorithm at last (Section 3.3).

3.1 Basic Style Transfer Framework

We adapt the objectives of self-reconstruction, back-translation, and style classification to our CBD framework. Let θ_E, θ_{D^l} , and θ_{D^r} denote the parameters of E, D^l, D^r , and $\theta = [\theta_E, \theta_{D^l}, \theta_{D^r}]$.

3.1.1 Self-Reconstruction

Self-reconstruction can warmly start the learning for non-parallel corpora and teach the model to preserve the content. Given an input x and its style s , if the target style $\tilde{s} = s$, the model is optimized to reconstruct x under both decoders, i.e., minimizing the self-reconstruction loss:

$$\begin{aligned} \mathcal{L}_{\text{rec}}(\theta) = & -\log D^l(x|E(x^*), s) \\ & -\log D^r(x|E(x^*), s) \end{aligned} \quad (1)$$

where x^* is a noisy version of x (by random word permutation and word removal) to avoid trivial solutions as in [Shen et al. \(2017\)](#).

3.1.2 Back-Translation

Through dynamically creating pseudo-parallel data, back-translation provides guidance for the transfer between different styles with increasing reliability as training proceeds. Given an input x and its style s , suppose we designate a target style $\tilde{s} \neq s$ and get $\tilde{x}^l \sim D^l(\tilde{x}|E(x), \tilde{s})$ and $\tilde{x}^r \sim D^r(\tilde{x}|E(x), \tilde{s})$. The model is optimized to restore x if we feed \tilde{x}^l or \tilde{x}^r as the input and s as the target style, i.e., minimizing the back-translation loss:

$$\mathcal{L}_{\text{back}}(\theta) = -\log D^l(x|E(\tilde{x}^l), s) - \log D^r(x|E(\tilde{x}^r), s) \quad (2)$$

This back-translation objective penalizes solutions that produce the same outputs for a given target style regardless of the inputs, thus alleviating the content deviation patterns in the OT problem.

3.1.3 Style Classification

Style classification enforces the style conversion goal by using a style classifier C (with parameters θ_C) to justify the style type of the transferred outputs. Given an input x and the target style \tilde{s} , suppose we get $\tilde{x}^l \sim D^l(\tilde{x}|E(x), \tilde{s})$ and $\tilde{x}^r \sim D^r(\tilde{x}|E(x), \tilde{s})$. The model is optimized to ensure \tilde{x}^l and \tilde{x}^r to be categorized to the style type \tilde{s} by C , i.e., minimizing the style classification loss:

$$\mathcal{L}_{\text{sty}}(\theta) = -\log C(\tilde{s}|\tilde{x}^l) - \log C(\tilde{s}|\tilde{x}^r) \quad (3)$$

3.2 Collaborative Learning

As discussed in Section 1, the OT / UT problems are actually the wrongly predicted words under the lack of ground-truth and the conflicting forces of the style conversion goal and content preservation goal. To provide more supervision, we establish a **mutual knowledge distillation** scheme between the two decoders. Since the two decoders are conditionally independent given the encoder’s outputs and the target style, we expect them to have inherently different knowledge acquisition processes. Then distilling the knowledge from one to the other can regularize each decoder by encouraging consistent predictions. Meanwhile, we explicitly ensure the two decoders’ inherent difference by a **distinguishability constraint** which employs a discriminator to distinguish their behaviors. Together with the opposite decoding direction, this constraint keeps the mutual knowledge distillation from rapidly pushing both decoders towards one bad local optimum where incorrect but consistent patterns are reinforced.

3.2.1 Mutual Knowledge Distillation

We regularize D^l and D^r via two-way knowledge distillation: both try to learn each other’s knowledge on producing the transferred output. Consider the knowledge distillation from D^r to D^l . Following the knowledge distillation framework (Hinton et al., 2015), given an input x and the target style \tilde{s} , D^l is optimized to decrease the KL divergence between its probability distribution over all possible outcomes with that of D^r , i.e., minimizing:

$$\mathcal{L}_{\text{mkd}}(\theta_{D^l}) = \text{KL}(D^r(\tilde{x}|E(x), \tilde{s}) \| D^l(\tilde{x}|E(x), \tilde{s}))$$

Eliminating the negative entropy term which is irrelevant to D^l from the KL divergence, $\mathcal{L}_{\text{mkd}}(\theta_{D^l})$ can be reformulated as

$$\sum_{t \in \mathcal{T}(x, \tilde{s})} -D^r(t|E(x), \tilde{s}) \log D^l(t|E(x), \tilde{s})$$

where $\mathcal{T}(x, \tilde{s})$ denotes all the possible transferred outcomes. However, exact computation for $\mathcal{L}_{\text{mkd}}(\theta_{D^l})$ is intractable with the summation over the exponential search space $\mathcal{T}(x, \tilde{s})$. Following Kim and Rush (2016), we approximate the target distribution $D^r(\tilde{x}|E(x), \tilde{s})$ as $\mathbb{1}[\tilde{x} = t^r]$, where $t^r = \arg \max_{t \in \mathcal{T}(x, \tilde{s})} D^r(t|E(x), \tilde{s})$ denotes the mode of the target distribution. As the maximization problem is still intractable, t^r is further approximated by a sequence t_*^r using greedy decoding or beam search on D^r . As a result, we arrive at

$$\mathcal{L}_{\text{mkd}}(\theta_{D^l}) = -\log D^l(t_*^r|E(x), \tilde{s}) \quad (4)$$

This resulted objective function is equivalent to optimizing D^l with pseudo-parallel data generated using D^r . Similarly, for the knowledge distillation from D^l to D^r , D^r is optimized to minimize:

$$\mathcal{L}_{\text{mkd}}(\theta_{D^r}) = -\log D^r(t_*^l|E(x), \tilde{s}) \quad (5)$$

3.2.2 The Distinguishability Constraint

The distinguishability constraint penalizes the cases where the two decoders lose their specialty in knowledge acquisition and collapse to each other. To this end, we jointly train a discriminator F (with parameters θ_F) to discriminate the behavior of the two decoders. Specifically, we represent the behavior of a decoder by the sequence of softmax probabilities associated with its transferred output.

Let $F(b)$ denote the probability of behavior b coming from D^l instead of D^r . Given an input x and the target style \tilde{s} , suppose we get

Algorithm 1 Training algorithm of CBD.

- 1: **Input:** non-parallel training corpus $\mathcal{M} = \{(x_i, s_i)\}_{i=1}^N$
 - 2: Initialize θ_C by pretraining a style classifier on \mathcal{M} ; initialize $\theta_E, \theta_{D^l}, \theta_{D^r}, \theta_F$ randomly
 - 3: **for** each iteration $j = 1, 2, \dots, L$ **do**
 - 4: Sample an text-style pair $(x, s) \sim \mathcal{M}$
 - 5: Sample a target style $\tilde{s} \sim \mathcal{S}$ with $\tilde{s} \neq s$
 - 6: Generate $\tilde{x}^l \sim D^l(\tilde{x}|E(x), \tilde{s})$
 - 7: Generate $\tilde{x}^r \sim D^r(\tilde{x}|E(x), \tilde{s})$
 - 8: Compute $\mathcal{L}(\theta, \theta_F)$ by Eq 7
 - 9: Update θ, θ_F by optimizing $\mathcal{L}(\theta, \theta_F)$
 - 10: Compute $\mathcal{L}_{sty-c}(\theta_C)$ by Eq 8
 - 11: Update θ_C by optimizing $\mathcal{L}_{sty-c}(\theta_C)$
 - 12: **end for**
-

$\tilde{x}^l \sim D^l(\tilde{x}|E(x), \tilde{s})$ and $\tilde{x}^r \sim D^r(\tilde{x}|E(x), \tilde{s})$, and $o(\tilde{x}^l)$ and $o(\tilde{x}^r)$ denote their softmax probability sequences. The decoders and F are optimized to ensure that $o(\tilde{x}^l)$ and $o(\tilde{x}^r)$ can be correctly classified by F , i.e., minimizing

$$\mathcal{L}_{\text{dis}}(\theta_{D^l}, \theta_{D^r}, \theta_F) = -\log F(o(\tilde{x}^l)) - \log(1 - F(o(\tilde{x}^r))) \quad (6)$$

Note that the distinguishability constraint is not incompatible with mutual knowledge distillation: while mutual knowledge distillation focuses on the consistency between the joint probabilities of two decoders, i.e., $D^l(\tilde{x}|E(x), \tilde{s})$ and $D^r(\tilde{x}|E(x), \tilde{s})$, the distinguishability constraint focuses on the difference between their factor sequences, i.e., $\{D^l(\tilde{x}_t|E(x), \tilde{s}, \tilde{x}_1^{t-1})\}_{t=1}^T$ and $\{D^r(\tilde{x}_t|E(x), \tilde{s}, \tilde{x}_{t+1}^T)\}_{t=1}^T$, where T denotes the sequence length.

3.3 Model Training

Integrating the collaborative learning mechanism with the basic framework, we formulate the full objective function for CBD as minimizing

$$\mathcal{L}(\theta, \theta_F) = \mathcal{L}_{\text{rec}}(\theta) + \mathcal{L}_{\text{back}}(\theta) + \alpha \mathcal{L}_{\text{sty}}(\theta) + \beta \mathcal{L}_{\text{mkd}}(\theta_{D^l}, \theta_{D^r}) + \gamma \mathcal{L}_{\text{dis}}(\theta_{D^l}, \theta_{D^r}, \theta_F) \quad (7)$$

where $\mathcal{L}_{\text{mkd}}(\theta_{D^l}, \theta_{D^r}) = \mathcal{L}_{\text{mkd}}(\theta_{D^l}) + \mathcal{L}_{\text{mkd}}(\theta_{D^r})$, and α, β and γ are hyperparameters.

The style classifier C is pretrained on \mathcal{M} and further updated in our training stage with the spirit of adversarial learning² (Goodfellow et al., 2014),

²The transfer model (incl. the encoder and the decoders), acting as the generator from the adversarial learning field,

which has been shown to stabilize the learning of CBD in our preliminary experiments. For an input x and its style s , we enforce C to correctly predict s as the style for x ; while for the outputs \tilde{x}^l and \tilde{x}^r produced by D^l and D^r under target style \tilde{s} , we enforce C to be uncertain between s and \tilde{s} by assigning a uniform distribution over the two styles (which represents the highest uncertainty). Formally, Eq 8 is minimized.

$$\mathcal{L}_{\text{sty-c}}(\theta_C) = - \sum_{\tilde{x} \in \{\tilde{x}^l, \tilde{x}^r\}} \sum_{s' \in \{s, \tilde{s}\}} \frac{1}{2} \log C(s'|\tilde{x}) - 2 \log C(s|x) \quad (8)$$

The training algorithm is summarized in Algorithm 1. The \tilde{x}^l and \tilde{x}^r in step 6 and 7 of Algorithm 1 are generated by greedy decoding, and they further act as t_*^l and t_*^r in Eq 5 and 4. Greedy decoding is also used during inference. Note that the discreteness of text generation hinders the gradient backpropagation from \mathcal{L}_{sty} to θ . We tackle this problem by approximating each discrete word with the softmax distribution given by the decoder.

4 Experiments

4.1 Experimental Settings

Datasets. We evaluate CBD on a sentiment transfer dataset YELP (Li et al., 2018) and a formality transfer dataset GYAFC (Rao and Tetreault, 2018). The YELP dataset is composed of business reviews from Yelp, with each review annotated as positive or negative. The GYAFC dataset is composed of sentences from Yahoo Answers, with each sentence annotated as formal or informal. Data statistics and preprocessing details are provided in Appendix A.

Implementation Details. The encoder and two decoders are implemented by single-layer Gated Recurrent Units (GRU) networks, while the style classifier and the discriminator employ the TextCNN architecture (Kim, 2014). In decoding, we follow Lample et al. (2019) and input the target style to the decoders as a special start token, which is then mapped to an embedding vector as the ordinary tokens. During inference, we produce two outputs for each sample (one from the D^l and the other from D^r) then select the output with larger log-probability (assigned by its origin

tries to produce a result making C predict its style as the given target style. With the first term in Eq 8, C acts as the discriminator from the adversarial learning field and is encouraged to be uncertain on the transferred results.

	YELP					GYAFC				
	ACC	BLEU	PPL	G2	H2	ACC	BLEU	PPL	G2	H2
CrossAligned (Shen et al., 2017)	74.9	9.1	43.5	26.1	16.2	67.7	3.6	25.8	15.6	6.8
StyleEmbedding (Fu et al., 2018)	8.4	21.1	44.1	13.3	12.0	24.5	7.9	60.4	13.9	11.9
MultiDecoder (Fu et al., 2018)	48.3	14.5	80.4	26.5	22.3	18.8	12.3	71.5	15.2	14.9
BackTrans (Prabhumoye et al., 2018)	95.4	2.5	19.7	15.4	4.9	65.6	0.9	57.2	7.7	1.8
CycledRL (Xu et al., 2018)	53.5	18.6	264.1	31.5	27.6	81.6	2.0	80.8	12.8	3.9
TemplateBased (Li et al., 2018)	84.9	22.6	181.7	43.8	35.7	52.1	35.2	87.5	42.8	42.0
RetrieveOnly (Li et al., 2018)	95.7	1.7	47.2	12.8	3.3	90.9	0.4	36.7	6.0	0.8
DeleteOnly (Li et al., 2018)	85.7	14.8	59.6	35.6	25.2	21.6	29.2	82.8	25.1	24.8
Del-Ret-Gen (Li et al., 2018)	89.7	16.0	56.3	37.9	27.2	50.3	21.2	69.2	32.7	29.8
UnsuperMT (Zhang et al., 2018c)	96.9	22.8	52.2	47.0	36.9	61.1	33.4	45.7	45.2	43.2
DualRL (Luo et al., 2019)	89.2	28.0	44.6	50.0	42.6	74.8	41.9	79.9	56.0	53.7
PointOperate (Wu et al., 2019a)	90.5	29.7	43.0	51.8	44.7	37.0	44.9	50.4	40.8	40.6
WordStyleRel (Zhou et al., 2020)	88.7	30.4	42.8	51.9	45.3	78.1	46.0	45.8	59.9	57.9
CBD	96.9	30.2	42.1	54.1	46.0	84.9	47.1	40.3	63.2	60.6

Table 2: Automatic evaluation results on the YELP dataset and the GYAFC dataset. G2: the geometric mean of ACC and BLEU. H2: the harmonic mean of ACC and BLEU.

decoder) as the final transferred result. We set $\alpha = 0.1$, $\beta = 0.1$ and $\gamma = 0.01$ in Eq 7. More details are provided in the Appendix B.

Baselines. We compare CBD with: (1) implicit disentanglement based methods including CrossAligned (Shen et al., 2017), StyleEmbedding (Fu et al., 2018), MultiDecoder (Fu et al., 2018), and BackTrans (Prabhumoye et al., 2018); (2) explicit disentanglement based methods including CycledRL (Xu et al., 2018), TemplateBased (Li et al., 2018), RetrieveOnly (Li et al., 2018), DeleteOnly (Li et al., 2018), and Del-Ret-Gen (Li et al., 2018); (3) non-disentanglement based methods including UnsuperMT (Zhang et al., 2018c), DualRL (Luo et al., 2019), PointOperate (Wu et al., 2019a), and WordStyleRel (Zhou et al., 2020).

4.2 Evaluation Measures

Following our baselines, we adopt both automatic evaluation and human evaluation to assess models on three aspects: style compatibility, content preservation, and fluency.

Automatic Evaluation. For style compatibility: A style classifier C_{eval} with the same architecture as C is independently learned on \mathcal{M} . We measure the style compatibility by the prediction accuracy (ACC) of C_{eval} on each model’s output, using the target styles as ground-truth labels. For content preservation: Each test sample has been associated with one human reference on YELP³ and

³As Luo et al. (2019) provides three additional references for each sample, we report BLEU scores based on this four-reference version in Appendix D.

four human references on GYAFC. We measure the content preservation by the BLEU score (using `multi-bleu.perl`⁴) between the model’s outputs and human references. For fluency: A language model LM with a single-layer GRU architecture is learned on all text sequences from \mathcal{M} . We measure the fluency by the Perplexity (PPL) of LM on the model’s outputs.

Human Evaluation. We invite three human annotators to evaluate different models’ outputs for 200 test samples on each dataset. The annotators score each transfer result from 1 (the lowest quality) to 5 (the highest quality) in terms of style compatibility, content preservation, and fluency. More details are provided in Appendix C.

4.3 Results and Analysis

Automatic Evaluation Results. Table 2 shows the automatic evaluation results on YELP and GYAFC. Overall, the non-disentanglement based methods demonstrate better performance than the implicit / explicit disentanglement based methods which tend to sacrifice content preservation for style compatibility, i.e., the OT problem. On YELP, our CBD performs the best on style compatibility, and is comparable to the best on content preservation. While it achieves second-best on fluency, the BackTrans model with the best fluency suffers from severe content loss with a low BLEU score. On GYAFC, our CBD performs the best on

⁴<https://github.com/moses-smt/mosesdecoder/blob/master/scripts/generic/multi-bleu.perl>

	YELP			GYAFC			U_{special}	U_{random}
	Style	Content	Fluency	Style	Content	Fluency		
CrossAligned (Shen et al., 2017)	3.6	2.5	3.2	1.6	1.3	3.5	21.8	15.4
Del-Ret-Gen (Li et al., 2018)	3.7	3.4	3.6	1.5	2.5	2.4	36.6	21.3
DualRL (Luo et al., 2019)	4.2	4.2	4.1	3.0	3.6	3.8	29.6	21.5
WordStyleRel (Zhou et al., 2020)	3.7	4.4	4.0	3.3	3.9	4.1	74.9	35.7
CBD	4.4	4.4	4.2	3.5	4.1	4.2	27.0	16.5

Table 3: **Middle blocks:** Human evaluation results on the YELP dataset and the GYAFC dataset. **Right block:** Percentage of UT results of different models on the YELP dataset, where a transferred result is regarded as an UT result if it still keeps the original sentiment parts.

	YELP: negative \rightarrow positive	GYAFC: formal \rightarrow informal
Input	prime rib was fatty and not cooked per requested .	approach her , say hello , and ask if she is single .
CrossAligned	prime rib was tender and really cooked us .	ask her and then if it is too long .
Del-Ret-Gen	prime rib was fatty , tender , flavorful and requested .	you her , you say her , and ask if she is you .
DualRL	prime rib was fatty and always cooked per requested .	approach her er
WordStyleRel	prime rib was delicious and not cooked per requested .	approach her , say hello , and ask if she is single
CBD	prime rib was good and cooked per requested .	approach her , say hi and ask if she s single

Table 4: Ouputs of different models on exemplary sentences from YELP and GYAFC.

content preservation and the second-best on style compatibility and the third on fluency. Still, the RetrieveOnly model with the best style compatibility and second-best fluency, and the CrossAligned model with the best fluency, are both limited on the remaining metrics. Table 2 also reports the geometric mean and harmonic mean of ACC and BLEU, i.e., the G2 and H2 scores, on which our CBD outperforms all the baselines. Hence, we conclude that CBD achieves a better balance on style compatibility, content preservation, and fluency.

Human Evaluation Results. The middle blocks of Table 3 show the human evaluation results on YELP and GYAFC. Due to the high evaluation cost, we only compare CBD with a subset of our baselines which achieve better balance on the three metrics for both datasets than other baselines in their category. On both datasets, our CBD achieves the best results on all three aspects. And consistent with the automatic results, the non-disentanglement based methods outperform the disentanglement based methods.

Qualitative Results. Table 4 shows the transfer results of different methods for exemplar sentences on YELP and GYAFC. We can see that CBD can produce fluent outputs, clearly expressing the target style without loss of other semantics. In contrast, other approaches present OT / UT problems or produce influent sentences. Specifically, the disentan-

glement based CrossAligned and Del-Ret-Gen are more prone to OT: the underlying semantics of *not cooked per requested* from the negative \rightarrow positive example on YELP and the meaning of the formal \rightarrow informal example on GYAFC are poorly preserved. The non-disentanglement based DualRL and WordStyleRel are more prone to UT: on YELP, the negative *fatty* or *not cooked per requested* are unchanged; on GYAFC, the changes are more limited than our CBD. More qualitative results and analysis are provided in Appendix E.

Discussions on the OT / UT Problems. Besides the qualitative results, models’ strengths towards the OT problem are indicated by the content preservation scores, i.e., BLEU in Table 2 and Content in Table 3. Thus we conclude CBD can alleviate the OT problem especially faced by disentanglement based methods. However, the UT problem is only partially indicated by the style compatibility scores, as failures on style conversion can also be caused by irrational modification on style indicators such as indicator removal. From the style compatibility scores, we can only conjecture CBD shows improvement over baselines on the UT problem. For better justification, we prepare two subsets of YELP: U_{special} containing 200 carefully selected samples with more than one style indicator (e.g., the input in Table 1), and U_{random} containing 200 random samples. Three human annotators are invited to label if a given transferred result has the

UT problem. The right block of Table 3 presents the ratio of UT cases for different models. All models have more UT cases on U_{special} , suggesting the UT problem occurs more often in inputs with more than one style indicator since partial transfer results can fool the style classifier. CBD outperforms all baselines except for CrossAligned. However, CrossAligned has serious OT problems by deviating the semantics to achieve the target style, which can be demonstrated by its content preservation scores from Tables 2 and Table 3.

Limitation. Despite the improvement over baselines, the UT problem is still quite challenging for our model compared to the OT problem, especially when the style is expressed in less frequent manners. For a negative \rightarrow positive sample on YELP: *they only received one star because you have to provide a rating.*, our model generates: *they received one star because you have to provide a great rating.*. This can be attributed to the lack of common-sense knowledge. More failure cases and analysis are provided in Appendix F.

4.4 Ablation Study

To better validate the effectiveness of the proposed CBD, we compare the following ablated variants:

- (1) L2R + $\{\mathcal{L}_{\text{basic}}\}$;
- (2) R2L + $\{\mathcal{L}_{\text{basic}}\}$;
- (3) L2R + R2L + $\{\mathcal{L}_{\text{basic}}\}$;
- (4) L2R + R2L + $\{\mathcal{L}_{\text{basic}}, \mathcal{L}_{\text{mkd}}\}$;
- (5) **L2R + R2L + $\{\mathcal{L}_{\text{basic}}, \mathcal{L}_{\text{mkd}}, \mathcal{L}_{\text{dis}}\}$**
- (6) L2R + L2R + $\{\mathcal{L}_{\text{basic}}, \mathcal{L}_{\text{mkd}}, \mathcal{L}_{\text{dis}}\}$;
- (7) R2L + R2L + $\{\mathcal{L}_{\text{basic}}, \mathcal{L}_{\text{mkd}}, \mathcal{L}_{\text{dis}}\}$

where the variant (5) corresponds to our CBD model, and $\mathcal{L}_{\text{basic}} = \{\mathcal{L}_{\text{rec}}, \mathcal{L}_{\text{back}}, \mathcal{L}_{\text{sty}}\}$.

Table 5 shows the automatic evaluation results of these variants on YELP. We have the following observations: First, the comparison between (3) and (1)/(2) shows that, shallow interactions by the shared encoder cannot give the two-decoder setting a clear advantage over the one-decoder setting. Second, the comparison between (4) and (3) shows that, mutual knowledge distillation can promote style compatibility and content preservation while sacrificing fluency a little. Third, the comparison between (5) and (4) shows that, involving the distinguishability constraint can achieve further improvement on all aspects. Fourth, the comparison between (5) and (6)/(7) shows that, settings

with two unidirectional decoders underperform the bidirectional setting for all aspects. We conclude that, with comparable fluency, CBD (variant (5)) is advantageous over the other variants in achieving both the style conversion goal and content preservation goal.

To provide a deeper insight, Table 5 presents the per-word entropy⁵ of each variant. The entropy measures the uncertainty of the model’s predictions. We can see that variants (4)(5) show lower per-word entropy values than (1)(2). As explained in Section 1, the single-decoder models can unconfidently struggle between new word generation and source word copying. However, the mutual knowledge distillation in (4)(5) provides additional supervision to the decoders by gradually reinforcing consistent patterns and thus improves their confidence in prediction. For unsupervised tasks, lower entropy values are preferred as it represents the model’s capability to filter the large proportion of wrong choices (Graça et al., 2009; Niu et al., 2012). Figure 1 in Appendix G illustrates the probability distributions of different variants when predicting specific words. Consistent with the per-word entropy values, CBD shows more peaked distributions than the single-decoder variants. Note that the per-word entropy value of (6) is higher than CBD while that of (7) is lower than CBD. This is possible: as the two unidirectional decoders have more close per-word distributions, if the distribution themselves are less peaked then the posterior distributions after mutual knowledge distillation can also be less peaked (as in (6)); on the other hand, if the distribution themselves are more peaked then so are the posterior distributions (as in (7)). This can be implied by the per-word entropy values of (1) and (2) where (1) has a higher entropy value than (2).

Computational Overhead. With an extra decoder and the introduced mutual knowledge distillation process plus the distinguishability constraint, on a single Nvidia’s GTX 1080Ti GPU, the training speed of CBD is about 0.5 times the single-decoder setting while the inference speed is about 0.55 times the single-decoder setting. However, as the two decoders make inference independently, the gap diminishes (the speed ratio is 0.9:1) when we clearly assign the two decoders to different CUDA

⁵For $\tilde{x} \sim D(\tilde{x}|E(x), \tilde{s})$ where D is the decoder, the per-word entropy refers to: $\frac{1}{|\tilde{x}|} \sum_{t=1}^{|\tilde{x}|} H(D(\cdot|E(x), \tilde{s}, \tilde{x}_1^{t-1}))$, where H is the Shannon Entropy. The per-word entropy value for a model is averaged on its outputs for all test samples.

No.	ACC	BLEU	PPL	Per-Word Entropy
(1)	92.5	25.5	41.6	0.8642
(2)	93.2	26.3	41.0	0.8389
(3)	93.2	26.6	40.8	0.8045
(4)	95.3*	29.6†	42.4	0.5509
(5)	96.9*	30.2†	42.1	0.5394
(6)	94.7*	27.2†	42.5	0.5846
(7)	94.9*	27.8†	43.0	0.5127

Table 5: Automatic evaluation results and per-word entropy values for different variants of our CBD on YELP. *(†): result significantly better than (1)(2) with $p < 0.1(0.05)$.

streams during inference. Multiple GPUs can also be utilized to parallelize the inference of the two decoders.

To summarize, as the style conversion goal and content preservation goal push the model towards conflicting directions when lacking ground-truth, single-decoder models try to focus on both directions and thus have high uncertainty in decoding and show OT or UT problems. By collaborative learning with two bidirectional decoders, the proposed CBD model breaks this uncertainty: it gives more direct guidance by reinforcing the consistent predictions of two distinguishable decoders, so that the higher-entropy predictions are redistributed towards the correct direction. As a result, the OT and UT problems are alleviated. See Appendix G for additional quantitative and qualitative results for different variants.

5 Conclusion

In this paper, we address the unsupervised text style transfer task from a novel multi-agent learning perspective. To overcome the over-transfer and under-transfer problems, we introduce a pair of collaborative bidirectional decoders. Our collaborative learning mechanism performs mutual knowledge distillation on the two decoders and guarantees the rationality of this distillation process by a distinguishability constraint together with their opposite decoding directions. Quantitative and qualitative results on two benchmark datasets validate the strength of our framework over various single-decoder baselines in achieving both the style conversion goal and the content preservation goal. Our code will be made publicly available at https://github.com/sunlight-ym/CBD_style_transfer.

Acknowledgements

We would like to thank all the reviewers and our group members for their constructive comments. The research work described in this paper has been supported by the Hong Kong Research Grants Council under the general research fund scheme (project number: PolyU 11204919).

References

- Tianchi Bi, Hao Xiong, Zhongjun He, Hua Wu, and Haifeng Wang. 2019. [Multi-agent learning for neural machine translation](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 856–865, Hong Kong, China. Association for Computational Linguistics.
- Avrim Blum and Tom Mitchell. 1998. Combining labeled and unlabeled data with co-training. In *Annual Conference Computational Learning Theory*.
- Ning Dai, Jianze Liang, Xipeng Qiu, and Xuanjing Huang. 2019. [Style transformer: Unpaired text style transfer without disentangled latent representation](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5997–6007, Florence, Italy. Association for Computational Linguistics.
- Yanai Elazar and Yoav Goldberg. 2018. [Adversarial removal of demographic attributes from text data](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 11–21, Brussels, Belgium. Association for Computational Linguistics.
- Zhenxin Fu, Xiaoye Tan, Nanyun Peng, Dongyan Zhao, and Rui Yan. 2018. [Style transfer in text: Exploration and evaluation](#). In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pages 663–670. AAAI Press.
- Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron C. Courville, and Yoshua Bengio. 2014. [Generative adversarial nets](#). In *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, pages 2672–2680.
- João Graça, Kuzman Ganchev, Ben Taskar, and Fernando C. N. Pereira. 2009. [Posterior vs parameter sparsity in latent variable models](#). In *Advances in Neural Information Processing Systems 22: 23rd Annual Conference on Neural Information Processing*

- Systems 2009. Proceedings of a meeting held 7-10 December 2009, Vancouver, British Columbia, Canada*, pages 664–672. Curran Associates, Inc.
- Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. 2015. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*.
- Zhiting Hu, Zichao Yang, Xiaodan Liang, Ruslan Salakhutdinov, and Eric P. Xing. 2017. **Toward controlled generation of text**. In *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*, volume 70 of *Proceedings of Machine Learning Research*, pages 1587–1596. PMLR.
- Vineet John, Lili Mou, Hareesh Bahuleyan, and Olga Vechtomova. 2019. Disentangled representation learning for text style transfer. In *Annual Meeting of the Association for Computational Linguistics*.
- Yoon Kim. 2014. **Convolutional neural networks for sentence classification**. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1746–1751, Doha, Qatar. Association for Computational Linguistics.
- Yoon Kim and Alexander M. Rush. 2016. **Sequence-level knowledge distillation**. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1317–1327, Austin, Texas. Association for Computational Linguistics.
- Diederik P. Kingma and Jimmy Ba. 2015. **Adam: A method for stochastic optimization**. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.
- Samuli Laine and Timo Aila. 2017. **Temporal ensembling for semi-supervised learning**. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net.
- Guillaume Lample, Sandeep Subramanian, Eric Michael Smith, Ludovic Denoyer, Marc’Aurelio Ranzato, and Y-Lan Boureau. 2019. **Multiple-attribute text rewriting**. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019*. OpenReview.net.
- Dong-Hyun Lee. 2013. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *International Conference on Machine Learning Workshop*.
- Juncen Li, Robin Jia, He He, and Percy Liang. 2018. **Delete, retrieve, generate: a simple approach to sentiment and style transfer**. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1865–1874, New Orleans, Louisiana. Association for Computational Linguistics.
- Fuli Luo, Peng Li, Jie Zhou, Pengcheng Yang, Baobao Chang, Xu Sun, and Zhifang Sui. 2019. **A dual reinforcement learning framework for unsupervised text style transfer**. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, pages 5116–5122. ijcai.org.
- Aman Madaan, Amrith Setlur, Tanmay Parekh, Barnabas Poczos, Graham Neubig, Yiming Yang, Ruslan Salakhutdinov, Alan W Black, and Shrimai Prabhumoye. 2020. **Politeness transfer: A tag and generate approach**. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1869–1881, Online. Association for Computational Linguistics.
- Gang Niu, Bo Dai, Makoto Yamada, and Masashi Sugiyama. 2012. **Information-theoretic semi-supervised metric learning via entropy regularization**. In *Proceedings of the 29th International Conference on Machine Learning, ICML 2012, Edinburgh, Scotland, UK, June 26 - July 1, 2012*. icml.cc / Omnipress.
- Kartikey Pant, Yash Verma, and Radhika Mamidi. 2020. Sentiinc: Incorporating sentiment information into sentiment transfer without parallel data. In *Advances in Information Retrieval*.
- Shrimai Prabhumoye, Yulia Tsvetkov, Ruslan Salakhutdinov, and Alan W Black. 2018. **Style transfer through back-translation**. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 866–876, Melbourne, Australia. Association for Computational Linguistics.
- Siyuan Qiao, Wei Shen, Zhishuai Zhang, Bo Wang, and Alan Yuille. 2018. Deep co-training for semi-supervised image recognition. In *European Conference on Computer Vision*.
- Ella Rabinovich, Raj Nath Patel, Shachar Mirkin, Lucia Specia, and Shuly Wintner. 2017. **Personalized machine translation: Preserving original author traits**. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 1074–1084, Valencia, Spain. Association for Computational Linguistics.
- Sudha Rao and Joel Tetreault. 2018. **Dear sir or madam, may I introduce the GYAFC dataset: Corpus, benchmarks and metrics for formality style transfer**. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 129–140, New Orleans, Louisiana. Association for Computational Linguistics.
- Shuo Ren, Zhirui Zhang, Shujie Liu, Ming Zhou, and Shuai Ma. 2019. **Unsupervised neural machine translation with SMT as posterior regularization**. In *The*

- Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*, pages 241–248. AAAI Press.
- Tianxiao Shen, Tao Lei, Regina Barzilay, and Tommi S. Jaakkola. 2017. [Style transfer from non-parallel text by cross-alignment](#). In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 6830–6841.
- Akhilesh Sudhakar, Bhargav Upadhyay, and Arjun Maheswaran. 2019. [“transforming” delete, retrieve, generate approach for controlled text style transfer](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3269–3279, Hong Kong, China. Association for Computational Linguistics.
- Chen Wu, Xuancheng Ren, Fuli Luo, and Xu Sun. 2019a. [A hierarchical reinforced sequence operation method for unsupervised text style transfer](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 4873–4883, Florence, Italy. Association for Computational Linguistics.
- Xing Wu, Tao Zhang, Liangjun Zang, Jizhong Han, and Songlin Hu. 2019b. [Mask and infill: Applying masked language model to sentiment transfer](#). In *International Joint Conference on Artificial Intelligence*.
- Jingjing Xu, Xu Sun, Qi Zeng, Xiaodong Zhang, Xuancheng Ren, Houfeng Wang, and Wenjie Li. 2018. [Unpaired sentiment-to-sentiment translation: A cycled reinforcement learning approach](#). In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 979–988, Melbourne, Australia. Association for Computational Linguistics.
- Zichao Yang, Zhiting Hu, Chris Dyer, Eric P. Xing, and Taylor Berg-Kirkpatrick. 2018. [Unsupervised text style transfer using language models as discriminators](#). In *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*, pages 7298–7309.
- Xiangwen Zhang, Jinsong Su, Yue Qin, Yang Liu, Rongrong Ji, and Hongji Wang. 2018a. [Asynchronous bidirectional decoding for neural machine translation](#). In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th Innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pages 5698–5705. AAAI Press.
- Yi Zhang, Tao Ge, and Xu Sun. 2020. [Parallel data augmentation for formality style transfer](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3221–3228, Online. Association for Computational Linguistics.
- Yi Zhang, Jingjing Xu, Pengcheng Yang, and Xu Sun. 2018b. [Learning sentiment memories for sentiment modification without parallel data](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1103–1108, Brussels, Belgium. Association for Computational Linguistics.
- Zhirui Zhang, Shuo Ren, Shujie Liu, Jianyong Wang, Peng Chen, Mu Li, Ming Zhou, and Enhong Chen. 2018c. [Style transfer as unsupervised machine translation](#). *arXiv preprint arXiv:1808.07894*.
- Zhirui Zhang, Shuangzhi Wu, Shujie Liu, Mu Li, Ming Zhou, and Tong Xu. 2019. [Regularizing neural machine translation by target-bidirectional agreement](#). In *Proceedings of the AAAI Conference on Artificial Intelligence*.
- Chulun Zhou, Liangyu Chen, Jiachen Liu, Xinyan Xiao, Jinsong Su, Sheng Guo, and Hua Wu. 2020. [Exploring contextual word-level style relevance for unsupervised style transfer](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7135–7144, Online. Association for Computational Linguistics.
- Long Zhou, Jiajun Zhang, and Chengqing Zong. 2019. [Synchronous bidirectional neural machine translation](#). *Transactions of the Association for Computational Linguistics*, 7:91–105.

A Dataset Details

We provide the statistics of the YELP dataset⁶ and the GYAFC dataset⁷ in Table 6. For the sentiment transfer dataset YELP, we use the same train/dev/test split from Li et al. (2018). For the formality transfer dataset GYAFC, we use the subset in the Family & Relationship domain. Although the dataset is a parallel corpus with aligned formal and informal pairs, following Luo et al. (2019); Zhou et al. (2020), we ignore the alignment information to cater to the unsupervised task. We use the same train/dev/test split from Luo et al. (2019); Zhou et al. (2020).

The YELP dataset has already been tokenized and lowercased. We tokenize and lowercase the

⁶<https://github.com/lijuncen/Sentiment-and-Style-Transfer/tree/master/data/yelp>

⁷<https://github.com/raosudha89/GYAFC-corpus>

sentences in GYAFC with `spacy`⁸. For both datasets, we construct a vocabulary to keep the 10K most frequent words in the dataset. Out-of-vocabulary words are mapped to a special token `<unk>`.

B Additional Implementation Details

The encoder adopts a single-layer bidirectional Gated Recurrent Units (GRU) network, with 256 hidden units in each direction. The L2R decoder and the R2L decoder both employ an attention-based single-layer unidirectional GRU network with 512 hidden units. The word embeddings are shared between the encoder and the two decoders, with a size of 128.

Our implementation is based on PyTorch (version 1.3.1) in Ubuntu 16.04. Models are trained on a single Nvidia’s GTX 1080Ti GPU with 11 Gbps GDDR5X memory. We use a batch size of 64 and train the model for 100K iterations. The Adam algorithm (Kingma and Ba, 2015) is utilized to optimize the model with a learning rate of 0.001.

The hyperparameters α , β , and γ in Eq 7 are tuned on the development set. Specifically, we search α , β , and γ over the values in {0.001, 0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1.0, 2.0}. Each value is evaluated based on three trials with different seeds which are integers uniformly sampled from [1, 999]. We have consistent observations on both datasets: (a) For α , {0.05, 0.1, 0.2} have similar performances, while smaller (larger) values will increase BLEU (ACC) but significantly decrease ACC (BLEU). (b) For β , the performance, especially BLEU, increases as β goes from 0.001 to 0.1. The ACC will be clearly degraded when $\beta = 0.2$. When $\beta \geq 0.5$, the model quickly produces empty outputs from both decoders as a trivial solution for mutual learning which seems to dominate the training. (c) For γ , the benefits are most significant when $\gamma = 0.01$. The model becomes unstable when further increasing γ , e.g., the outputs may become quite influent with repeated tokens. As a result, we set $\alpha = 0.1$, $\beta = 0.1$, and $\gamma = 0.01$.

Following Lample et al. (2019), the gradients of the back-translation loss (Eq 2) will not be back-propagated to the generation pass for \tilde{x}^l and \tilde{x}^r .

C Human Evaluation Details

Each of the three human annotators fulfills the following requirements: (a) the annotator is well-

⁸<https://spacy.io/>

Dataset	Style	Train	Dev	Test
YELP	Positive	270K	2000	500
	Negative	180K	2000	500
GYAFC	Formal	51K	2247	500
	Informal	51K	2788	500

Table 6: Dataset statistics.

Models	BLEU
CrossAligned (Shen et al., 2017)	17.9
StyleEmbedding (Fu et al., 2018)	42.3
MultiDecoder (Fu et al., 2018)	27.9
BackTrans (Prabhumoye et al., 2018)	5.0
CycledRL (Xu et al., 2018)	37.0
TemplateBased (Li et al., 2018)	45.5
RetrieveOnly (Li et al., 2018)	2.9
DeleteOnly (Li et al., 2018)	29.0
Del-Ret-Gen (Li et al., 2018)	31.1
UnsuperMT (Zhang et al., 2018c)	44.5
DualRL (Luo et al., 2019)	55.2
PointOperate (Wu et al., 2019a)	59.2
WordStyleRel (Zhou et al., 2020)	60.4
CBD	59.7

Table 7: BLEU scores (evaluated with four references for each sample) of different models on the YELP dataset.

educated on English linguistics; (b) the annotator uses English social media sites frequently; and (c) the annotator is not one of the authors.

Following Li et al. (2018), for each test sample and the target style, each annotator was shown the outputs of all evaluated models. Different models’ outputs are randomly permuted. Before evaluation, for each dataset and each transfer direction, annotators are trained by: (a) instructions on the desirable properties of the text style transfer task; (b) the detailed interpretation for each level (1-5) of the three aspects: style compatibility, content preservation, and fluency; and (c) four exemplary transfer outputs on a source sentence associated with scores assigned by authors and a short explanation for the scores.

We measure the inter-annotator consistency of the human evaluation results by the Fleiss’ kappa score. Specifically, the Fleiss’ kappa score is 0.782 on YELP and 0.791 on GYAFC.

D BLEU Scores Using Four References on YELP

Table 7 provides the BLEU scores of different models calculated with four references for each test sample (provided by Luo et al. (2019)) on YELP.

Consistent with the results in Table 2, our CBD is only slightly worse than the WordStyleRel model and outperforms all other baselines.

E Additional Qualitative Results

To better illustrate the improvement of our CBD over the baselines against the over-transfer and the under-transfer problems, we present additional qualitative examples from YELP and GYAFC in Table 8 and Table 9, respectively. The results show consistent patterns with those in Table 4. Specifically, the disentanglement based methods, especially the implicit disentanglement based CrossAligned model, suffer from serious over-transfer problem by losing original content or adding new content; on the other hand, the non-disentanglement based baselines tend to under-transfer by keeping part of the original sentiment semantics on YELP and by making limited transformations on GYAFC. In contrast, our CBD demonstrates better robustness towards both the over-transfer and the under-transfer problems.

F Failure Cases

We also present two more undesirable transfer results of our CBD for each dataset in Table 10. These failed cases mainly under-transfer the source sentences, showing the imperfection of our model towards the under-transfer problem in the following situations: (a) the style information is expressed in less frequent manners (compared to those indicated by adjectives) such as “with lots to see and try”; (b) the style information is expressed by words which can represent different styles in different context: for example, “hot” has been used to indicate both positive and negative sentiment in the training corpus; (c) partial changes can also be regarded as reasonable results for inherently continuous style types such as the formality transfer: for the formal \rightarrow informal example on GYAFC, the outputs of our CBD are still limited by only removing the comma, while more changes like changing “you” to “u” or removing the period can be further applied to achieve a more informal style. Besides, there are incorrect transfer results such as changing “ur” to “your” instead of “you are” for the informal \rightarrow formal example, which might be plausible for some cases while not plausible for the given context. Based on the above observations, we conclude the limitations of our CBD include: first, it cannot fully utilize the structure and/or context of

the source sentence to make the transfer; second, it cannot control how much style information is transferred for inherently continuous style types. We leave exploration for these issues in future work. In this paper, we focus on problems brought by the conflicting driving forces of the style conversion goal and the content preservation goal.

G Detailed Ablation Study

In this section, we provide more details for our ablation study on the YELP dataset to validate the effectiveness of the CBD model.

G.1 Qualitative Results of Different Ablated Variants

To provide more insights into different ablated variants, Table 11 demonstrates the transferred results of variants (1)-(7) on eight samples from YELP. We can observe that the single-decoder variants (1) and (2) easily suffer from over-transfer (e.g., losing the “walmart” in the third negative \rightarrow positive sample) or under-transfer problems (e.g., keeping “was even better” in the last positive \rightarrow negative sample). The variant (3) can only perform better for some cases by setting up a shallow connected two-decoder scheme. With mutual knowledge distillation, variant (4) is much less prone to the over-transfer and under-transfer problems. However, it can emphasize the consistency between the decoders too strongly and may still lead to suboptimal results (e.g., totally under-transfer the last negative \rightarrow positive sample). By incorporating a distinguishability constraint, our CBD, i.e., variant (5), can alleviate both problems. In contrast, (6) and (7) with two unidirectional decoders, perform even worse than (4) for most cases. This further implies that the inherent difference is quite limited for unidirectional decoders, therefore, the two decoders may have similar bad patterns which are further reinforced during training: take the third positive \rightarrow negative case for example, both (1) and (6) suffer from the over-transfer problem while both (2) and (7) suffer from the under-transfer problem.

As shown in Table 5, the per-word entropy values of two-decoder settings are lower than those of single-decoder settings. To better illustrate this, Figure 1 presents the top-5 predicted words together with their probabilities of variants (1), (2), (5), and (7) when they predict (a) the word after “would” and (b) the word after “dentistry” given the third positive \rightarrow negative input from Table 11.

	YELP: negative → positive	YELP: positive → negative
Input	it was over fried and very hard .	my dr pepper ribs were excellent and very tender .
CrossAligned	it was very tasty and very hard .	my husband ordered bacon were chicken and very greasy and mushy .
Del-Ret-Gen	the food was very good fried and very hard .	my dr pepper ribs left very tender .
DualRL	it was cooked fried and very sweet .	my dr pepper ribs were mediocre and very tender .
WordStyleRel	it was over fried and very good .	my dr pepper ribs were bland and very tender .
CBD	it was perfectly fried and very good .	my dr pepper ribs were terrible and very tasteless .
	YELP: negative → positive	YELP: positive → negative
Input	always rude in their tone and always have shitty customer service !	i 'm a regular drive-through customer at this location .
CrossAligned	always authentic all other and and they are the food !	i 'm a better fan and at this location for possible .
Del-Ret-Gen	i always enjoy going in always their kristen and always have shitty customer service !	horrible experience i 'm a regular drive-through customer at this location never again .
DualRL	always friendly in their best price and always have loved customer service !	i 'm a regular regular customer at this location .
WordStyleRel	always friendly in their tone and always have the customer service !	i 'm a horrible drive-through customer at this location .
CBD	always nice in their tone and always have great customer service !	i 'm not a regular drive-through customer at this location .
	YELP: negative → positive	YELP: positive → negative
Input	so , no treatment and no medication to help me deal with my condition .	the service was great and would gladly go back .
CrossAligned	so , her and and my hair and gave me up and happy .	the service was better and would never go back back .
Del-Ret-Gen	so service was great , no treatment and no to help me with no treatment .	would gladly anyone back go back .
DualRL	so , great treatment and great help me deal with my condition .	the service was horrible and would gladly go back .
WordStyleRel	so , great treatment and no medication to help me deal with my condition .	the service was horrible and would gladly go back .
CBD	so , great treatment and great medication to help me deal with my condition .	the service was horrible and would not go back .

Table 8: Additional results of different models on exemplary sentences from YELP.

	GYAFC: formal → informal	GYAFC: informal → formal
Input	make sure it is your decision .	do nt try to figure it out that s jus the way it is .
CrossAligned	think it is not a friend .	try to tell him and it is not the same thing too .
Del-Ret-Gen	make sure your not it is your decision .	do not you try to you it out thats you the way it is .
DualRL	make sure it is your decision er	do not try to figure it out the way it is .
WordStyleRel	make sure it is your life	simply try to figure it out . a the way it is .
CBD	make sure its your decision	do not try to figure it out , that is simply the way it is .
Input	do not approach her and let her know that you find her looks very attractive .	its all about her , and no it should nt be something that happends on a first date .
CrossAligned	just tell him if you dont get out with him and dont just be happy	it is possible and that is not a good thing and she is a good man .
Del-Ret-Gen	approach her and let her know that you find her looks ! !	i would like her to call , and everyone has something that you on a first date .
DualRL	dont approach her and let her know that you find her looks very screwed ...	it is all about her , and no it be something that upsets a first .
WordStyleRel	do n't approach her and let her know that you find her looks very attractive	“ all about her , and no it shouldnt be something that catches on a first date .
CBD	do nt approach her and let her know that u find her looks very hot	it is all about her , and no , it should not be something that you have experienced on a first date .
Input	it is a turn off if she sits on your best friend's lap when drunk .	no u should nt leave them... just teach them what to do to please u better...
CrossAligned	it depends on a good thing if she looks cheat on a	i need to send her and do not know what you like to do .
Del-Ret-Gen	you off if she you on your best friend's you when you 't you.	no, u does not leave u .
DualRL	im a turn off if she sits on your best friend's understanding when she 's drunk .	no, you leave them just teach them what to do.
WordStyleRel	it is a turn off if she sits on your best friend's lap when drunk	perhaps you must leave them. just teach them what to do to please you better.
CBD	its a turn off if she sits on ur best friend's lap when drunk	no, you should not leave them. just teach them what to do to please you better.

Table 9: Additional results of different models on exemplary sentences from GYAFC.

	YELP: negative → positive	YELP: positive → negative
Input	but it was disgusting and hot in there .	it is a cool place , with lots to see and try .
CrossAligned	but it was attentive and nice every time .	it is a long place , run down to fix to use home .
Del-Ret-Gen	the food is good but they have the best hot in there .	it is my waste of time , with lots to try and see .
DualRL	but it was delicious and hot in there .	it is a frustrating place , with lots to see and try .
WordStyleRel	but it was delicious and hot in there .	it is a horrible place , with lots to see and try .
CBD	but it was fantastic and hot in there .	it is a depressing place , with lots to see and try .
	GYAFC: formal → informal	GYAFC: informal → formal
Input	if a man cares about you , then he will call .	it all depends on when ur ready .
CrossAligned	if your a girl , then then then she will not .	it depends on what type of you ?
Del-Ret-Gen	if a man you is about it ... you will call .	it all depends on when ur open .
DualRL	if a man cares about you er	it all depends on when ready .
WordStyleRel	if a man cares about you , then he will call u	it all depends on when your ready .
CBD	if a man cares about you then he will call .	it all depends on when your ready .

Table 10: Failure cases of our CBD on YELP and GYAFC. We also include the results of our baselines for comparison.

	YELP: negative → positive	YELP: positive → negative
Input	sketchy sketchy sketchy pizza delivery !	the food is great here and very authentic .
(1)	diverse diverse delivery and the pizza !	the food is n't here very authentic .
(2)	cute sketchy pizza spot delivery !	i was cold here and nearly frozen .
(3)	cool cool pizza delivery !	the food is horrible here and very overpriced .
(4)	impressive impressive pizza delivery !	the food is horrible here and very sloppy .
(5)	cute cute cute pizza delivery !	the food is horrible here and very bland .
(6)	amazing sketchy pizza delivery !	the food is n't here very authentic .
(7)	amazing amazing amazing pizza delivery !	sat down here and @num degrees .
Input	i have not received such an attitude toward a customer before .	the food is good and from what i can tell is rather authentic .
(1)	i have always received such an amazing customer service before .	the food is not what from what i can tell is rather overpriced .
(2)	i have always received such a team toward a customer before .	however , and from what i can tell was rather frozen .
(3)	i have received such an excellent customer vibe toward a customer before .	the food is not from what i can tell is rather overpriced .
(4)	i have always received such happy hour toward a customer before .	too bad and from what i can tell is rather sloppy .
(5)	i have always received such an excellent attitude toward a customer before .	the food is bad and from what i can tell is rather bland .
(6)	i have always received such an exceptional attitude toward a customer before .	the food is not what i can tell from is rather authentic .
(7)	i have always received such fantastic toward a customer before .	@num and from what i can tell is not authentic .
Input	this is the worst walmart neighborhood market out of any of them .	i would recommend dentistry of old town scottsdale to everyone .
(1)	this is the best neighborhood neighborhood market out of them .	i would not recommend any old dentistry to town scottsdale to everyone .
(2)	this is the best craft neighborhood market out of any of them .	i would recommend dentistry of old town of scottsdale tonight .
(3)	this is my favorite neighborhood market out of any of them .	@num yr old dentistry of old scottsdale to anyone .
(4)	this is the best neighborhood market out of any of them .	i would not recommend dentistry of old town scottsdale to everyone .
(5)	this is the best walmart neighborhood market out of any of them .	i would not recommend dentistry of old town scottsdale to everyone .
(6)	this is the best neighborhood market out of any of them .	i would not recommend old town or scottsdale to anyone .
(7)	this is the best neighborhood market out of any of them .	i would suggest dentistry of old town scottsdale to everyone .
Input	bottom line they over promise and under deliver .	well the food was great and the price of it was even better .
(1)	bottom line over they promise and under deliver .	apparently the food was n't great and the price of it was even better .
(2)	bottom line they over promise and deliver .	well the food was cold and the price of it was even better .
(3)	top line they are over promise and they deliver .	well the food was horrible and the price of it was even better .
(4)	bottom line they over promise and under deliver .	well the food was horrible and the price of it was even better .
(5)	bottom line they promise and perfectly deliver .	well the food was horrible and the price of it was not even better .
(6)	bottom line they cute promise and under deliver .	well the food was n't the price of it was even worse .
(7)	bottom line well they promise and under deliver .	well the food was horrible and the price of it was even better .

Table 11: Results of different ablated variants on exemplary sentences from YELP.

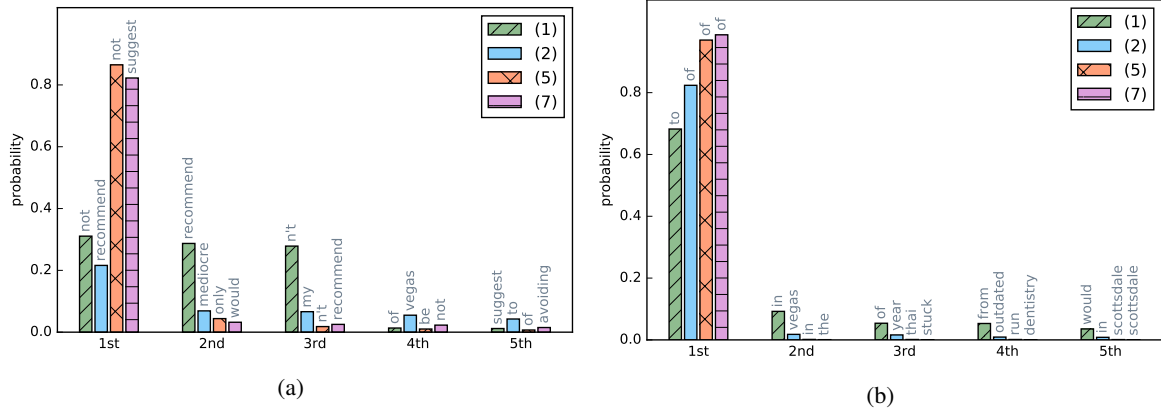


Figure 1: The top-5 words’ probabilities of variants (1), (2), (5) and (7) when they predict (a) the word after “would” and (b) the word after “dentistry” given the third positive \rightarrow negative input from Table 11.

No.	Style	Content	Fluency	U_{special}	U_{random}	$\text{BLEU}(D^l, D^r)$	F_{acc}
(4)	4.3	4.3	4.2	29.2	18.0	72.6	N/A
(5) CBD	4.4	4.4	4.2	27.0	16.5	72.4	55.2

Table 12: **Second block:** Human evaluation results of variants (4) and (5) on YELP. **Third block:** Percentage of under-transfer results of variants (4) and (5) on YELP. **Fourth block:** BLEU scores between the L2R decoder and R2L decoder of variants (4) and (5), and the accuracy of the discriminator of variant (5) on YELP.

Variants (5) and (7) demonstrate more peaked distributions over the words, which conforms to the lower entropy values. However, the most probable word of variant (7) in Figure 1a, i.e., “suggest”, is an incorrect prediction expressing the opposite of the target style. This further shows two unidirectional decoders may amplify the bad patterns learned by the inherently similar decoders.

G.2 Effect of the Distinguishability Constraint

Based on the quantitative and qualitative results, variant (4) without the distinguishability constraint shows stronger performance which is quite close to CBD (variant (5)) than other ablated variants. To better explore the effect of distinguishability constraint, we incorporate variant (4) in human evaluation. As shown in Table 12, CBD can improve variant (4) on style compatibility (partially reflecting the under-transfer problem) and content preservation (reflecting the over-transfer problem). Furthermore, Table 12 also reports their percentages of under-transfer cases on U_{special} and U_{random} . Still, CBD achieves a clear advantage. Hence, we conclude that incorporation of the distinguishability constraint leads to better capabilities, though limited, to address the over-transfer and the under-transfer problems.

Another question is how divergent the two de-

coders are at the end of training. Table 12 presents the BLEU scores between the outputs of the L2R decoder and the R2L decoder for variant (4) and CBD. We can observe that the BLEU scores of the two variants are comparable, both exhibiting a high similarity. Moreover, the discriminator cannot well distinguish behaviors of the two decoders. There is no wonder for this as the other loss functions dominate the learning process. Obviously, we can decrease the BLEU score and increase the accuracy of the discriminator by assigning a larger value to γ , i.e., the weight of the distinguishability constraint. However, this focuses on the wrong point and can only lead to worse results where influence from other objectives gets weakened and at least one decoder tends to produce unreasonable outputs to maximize their difference. Our explanation is that the distinguishability constraint behaves as an assistant or a regularizer for mutual knowledge distillation. While it is less significant for the performance than the mutual learning, it constrains the model to update in a more cautious way to avoid their collapsing and reinforcing incorrect but consistent patterns (e.g., keeping “was even better” in the last positive \rightarrow negative sample in Table 11).