Matching Structure for Dual Learning

Hao Fei[1], Shengqiong Wu[1], Yafeng Ren[2], Meishan Zhang[3]
1. Sea-NExT Joint Lab, National University of Singapore, Singapore
2. Guangdong University of Foreign Studies, Guangzhou, China
3. Harbin Institute of Technology, Shenzhen, China

Introduction

Many NLP tasks come in dual forms, such as neural machine translation (NMT), paraphrase generation, image captioning vs. text-to-image generation, text classification vs. conditioned text generation, semantic parsing vs. language generation, etc. Dual learning therefore has been proposed to model the duality between the primal and dual tasks, by minimizing the gap between joint distributions of the two tasks respectively. Formally, a dual learning system comprises 1) a primal task that maps $x \in \mathcal{X}$ to $y \in \mathcal{Y}$, i.e., $f_\theta : x \mapsto y$; and 2) a dual task mapping $y \in \mathcal{Y}$ to $x \in \mathcal{X}$, i.e., $g_\phi : y \mapsto x$.

However, we notice that the current dual learning scheme fails to explicitly model the structure correspondence between two coupled tasks. The integration of structure knowledge has been extensively exploited for enhancing the feature learning in a wide range of NLP tasks, which offers additional bias from a lower-level perspective (e.g., syntactic or linguistic) for better task-semantic inference. Unfortunately, the study of structure integration for dual learning has left unexplored. Given a pair of task, not only do they share the same input and output (in reverse), but it is often a close correspondence of the intermediate structures between them.

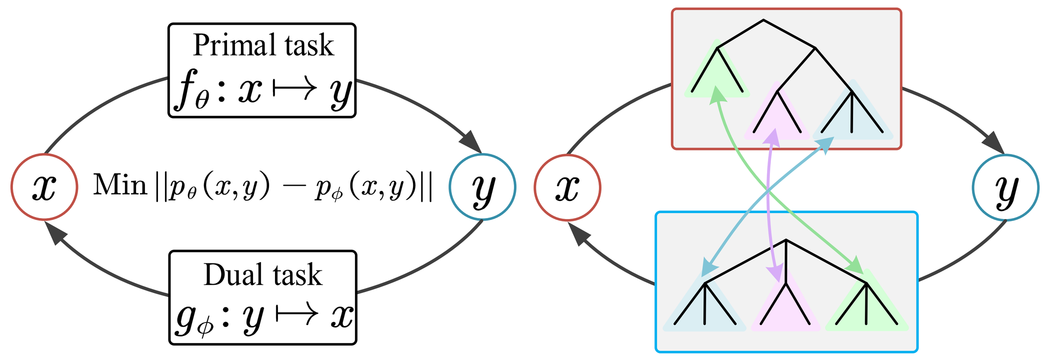

Figure 1: Left: dual learning framework. Right: dual learning with alignment of structural supervision.

To close the gap, this paper proposes matching the structure for dual learning. As shown in Figure 1, based on the vanilla dual learning framework, we perform structural alignment unsupervisedly between the primal and dual tasks, bridging them with structure connections.

Dually-Syntactic Structure Matching

Dually-Syntactic Structure Encoding. The input for both the primal and dual task is the sentential words $\{w_1, \cdots, w_n\}$. Meanwhile we have its syntactic constituency parse $\mathcal{T} = \{T_k\}_{k=1}^{K}$, where T_i is an intermediate constituency phrase or terminal word, and K denotes the total node number. Here we take the N-ary TreeLSTM as the structure encoder.

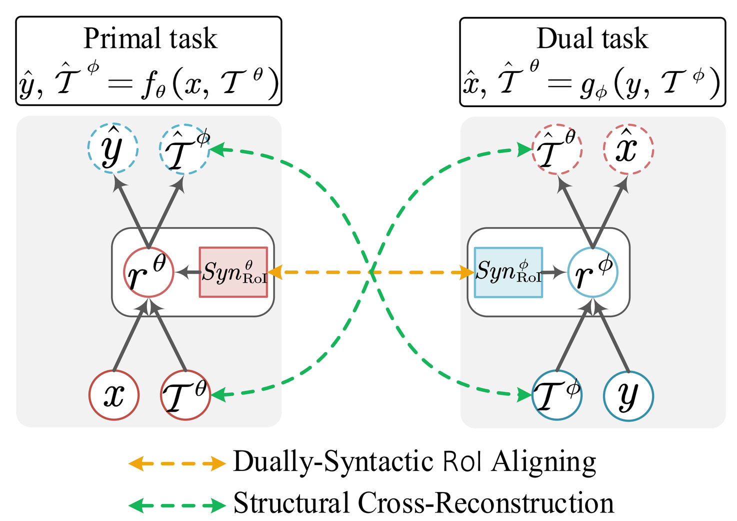

Figure 2: Symmetrically syntactic structure matching for dual learning.

Syntactic RoI Alignment. The core idea is to build the fine-grained structure correspondences between primal and dual tasks, pushing those pairs that serve the similar role in the context to be closer, i.e, $p(T_i|\mathcal{T}^\theta) \approx p(T_j|\mathcal{T}^\phi)$. Specifically,

$$p(T_i|\mathcal{T}) = \mathrm{Sigmoid}(\mathrm{FFNs}(\mathrm{Att}(T_i|\mathcal{T}))) \, .$$

Contrastive Region Repelling. We use the contrastive representation learning for the automatic structure matching.

Structural Cross-Reconstruction. On the other hand, during the text generation of $\hat{y}$ we make the model meanwhile to reproduce the corresponding syntax tree structure $\hat{\mathcal{T}}^\theta$. The syntax structure of the inut text from the opposite side (i.e, $\mathcal{T}^\theta$) can serve as a supervised signal. The benefits of such structural cross-reconstruction are multiple: making the structural awareness in the dual modeling more sufficient, providing additional syntactic constraint for the procedure, and also ensuring a global view during the generation.

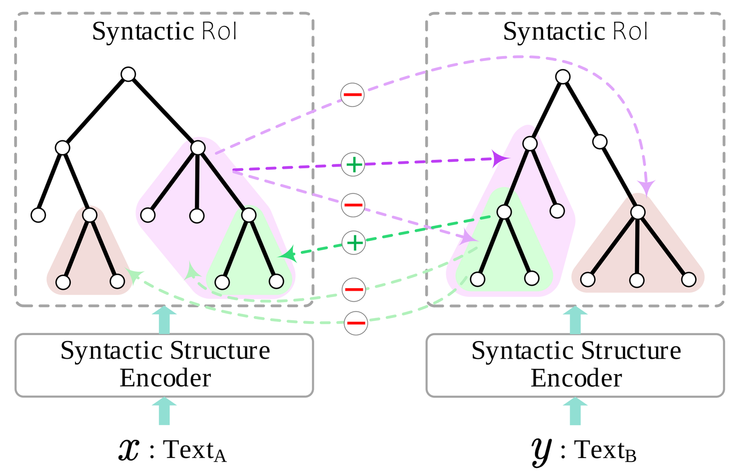

Figure 3: Dually-syntactic RoI alignment.

Exp-I: Text↔Text Applications. We examine the usefulness of the dually-syntactic structure matching for text↔text dual learning.

		ParaNMT						QUORA					
		B	R-1	R-2	R-L	B	R-1	B	R-1	R-2	R-L	B	R-1
• *Baseline*	B1	20.4	50.3	25.2	51.6	21.8	46.4	19.5	40.6	22.5	44.6	17.8	44.1
	B2	20.8	49.6	28.4	48.6	19.0	45.0	22.3	56.4	26.2	52.3	21.0	52.8
	B3	23.6	54.8	32.0	58.3	25.4	48.7	30.4	62.6	42.7	65.4	28.1	60.5
	B4	27.5	60.6	36.9	54.5	27.2	53.2	35.8	68.1	45.7	70.2	35.6	65.7
• *Transformer-based*	M1	24.6	50.3	30.7	45.8	25.4	51.7	29.7	58.5	37.5	59.6	28.0	60.5
	M2	27.2	56.4	34.4	50.6	26.1	53.6	33.4	63.4	41.8	63.4	34.8	65.8
	M3	26.2	57.1	33.0	53.5	27.8	55.9	32.0	65.7	40.0	66.4	34.0	64.3
	M4(RANK)	30.1	61.8	38.9	59.8	30.2	62.5	37.3	70.4	47.2	72.4	37.4	71.2
	M4(CL)	30.5	62.4	39.4	60.4	30.6	62.7	37.5	70.5	47.6	72.5	37.5	71.5
	OnlySyn	27.7	58.9	34.9	54.7	28.0	56.2	33.7	66.4	42.0	67.1	35.0	65.8
	-SAln	28.0	59.6	35.8	56.0	28.6	57.3	34.6	67.6	43.2	68.9	35.8	67.4
	-SyRec	29.7	60.2	37.8	58.3	29.7	61.0	36.1	68.9	45.0	71.4	36.5	69.3
	M3+BART	33.8	65.7	41.8	62.8	32.7	64.0	41.5	73.3	49.4	74.2	42.0	71.5
	M4+BART	36.7	66.2	43.6	64.0	34.8	64.6	43.0	74.8	52.8	76.8	43.5	72.8

Table 2. Results on paraphrase generation (SRC→TGT, SRC←TGT). B: BLEU, R-X: ROUGE-X.

		ParaNMT						QUORA					
		B	R-1	R-2	R-L	B	R-1	B	R-1	R-2	R-L	B	R-1
• *Baseline*	B1	20.4	50.3	25.2	51.6	21.8	46.4	19.5	40.6	22.5	44.6	17.8	44.1
	B2	20.8	49.6	28.4	48.6	19.0	45.0	22.3	56.4	26.2	52.3	21.0	52.8
	B3	23.6	54.8	32.0	58.3	25.4	48.7	30.4	62.6	42.7	65.4	28.1	60.5
	B4	27.5	60.6	36.9	54.5	27.2	53.2	35.8	68.1	45.7	70.2	35.6	65.7
• *Transformer-based*	M1	24.6	50.3	30.7	45.8	25.4	51.7	29.7	58.5	37.5	59.6	28.0	60.5
	M2	27.2	56.4	34.4	50.6	26.1	53.6	33.4	63.4	41.8	63.4	34.8	65.8
	M3	26.2	57.1	33.0	53.5	27.8	55.9	32.0	65.7	40.0	66.4	34.0	64.3
	M4(RANK)	30.1	61.8	38.9	59.8	30.2	62.5	37.3	70.4	47.2	72.4	37.4	71.2
	M4(CL)	30.5	62.4	39.4	60.4	30.6	62.7	37.5	70.5	47.6	72.5	37.5	71.5
	OnlySyn	27.7	58.9	34.9	54.7	28.0	56.2	33.7	66.4	42.0	67.1	35.0	65.8
	-SAln	28.0	59.6	35.8	56.0	28.6	57.3	34.6	67.6	43.2	68.9	35.8	67.4
	-SyRec	29.7	60.2	37.8	58.3	29.7	61.0	36.1	68.9	45.0	71.4	36.5	69.3
	M3+BART	33.8	65.7	41.8	62.8	32.7	64.0	41.5	73.3	49.4	74.2	42.0	71.5
	M4+BART	36.7	66.2	43.6	64.0	34.8	64.6	43.0	74.8	52.8	76.8	43.5	72.8

Table 2. Results on paraphrase generation (SRC→TGT, SRC←TGT). B: BLEU, R-X: ROUGE-X.

Syntactic-Semantic Structure Matching

It can be a broader range of NLP scenarios with dual learning technique where the task pair often includes non-text modalities, such as labels, image or audio etc. This makes the structure matching idea for text↔non-text dual learning non-trivial. We extends the above method of dually-syntactic structure matching to a method of syntactic-semantic structure matching. Since the task of non-text modality comes without explicit syntactic structure, our main idea is to take the semantic structure of non-text, and perform syntactic-semantic RoI alignment instead. Meanwhile, the syntactic structure reconstruction for the global-level benefit becomes structural unilateral-reconstruction.

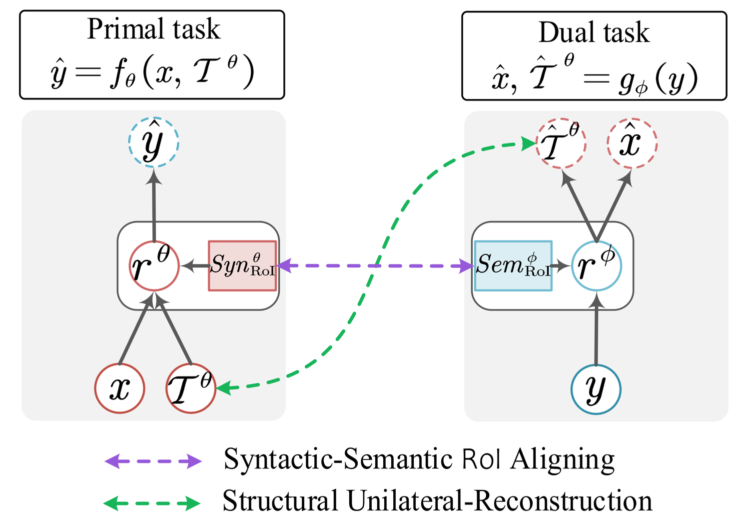

Figure 4: Syntactic-semantic structure matching.

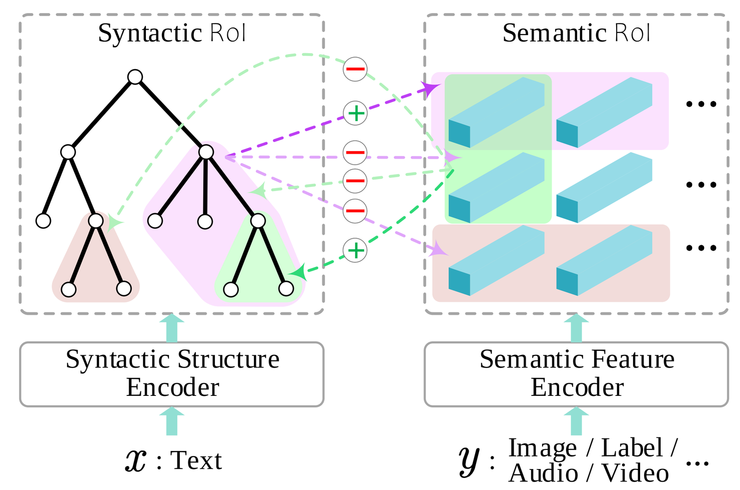

Figure 5: Syntactic-semantic RoI alignment via contrastive representation learning.

Exp-II: Text↔Non-Text Applications. Here we present the evaluations of our method in this section for text↔non-text scenarios.

	MsCoCo				Flickr30k			
	IS↑	FID↓	B-4	MTR	IS↑	FID↓	B-4	MTR
M1	25.6	28.3	32.5	22.8	6.8	36.8	17.6	15.5
M2	27.8	25.5	/	/	7.5	35.0	/	/
M3	28.4	24.8	36.1	25.1	7.3	34.2	20.1	17.2
M4	30.7	20.6	40.0	29.6	8.0	30.9	22.6	19.5
-SAln	29.0	21.5	37.3	28.3	7.4	33.0	21.3	17.9
-SyRec	29.8	21.3	39.2	29.0	7.7	31.8	21.9	18.6

Table 3. Results on text↔image experiment (TXT→IMG: text-to-image synthesis, TXT←IMG: image captioning). B-4: BLEU-4, MTR: METEOR.

	Yelp2014				IMDB			
	ACC	B-4	MTR	ACC	ACC	B-4	MTR	ACC
M1	60.6	17.8	33.0	53.8	50.6	17.6	36.9	43.6
M2	61.8	/	/	/	51.9	/	/	/
M3	62.0	19.4	36.4	56.6	53.8	18.3	41.4	47.3
M4	63.8	21.8	40.8	62.4	55.6	20.2	47.1	50.9
-SAln	63.2	19.9	37.0	57.2	54.2	18.9	44.6	48.4
-SyRec	62.9	20.4	38.5	61.8	55.0	19.5	46.0	49.3

Table 4. Results on Text↔Label experiment (TXT→LB: text classification, TXT←LB: conditioned text generation).

Analysis and Discussion

Here we further explore the underlying mechanisms how the structure matching improves.

Evaluating Correctness of Unsupervised Structure Matching. ★ First, structure matching helps correctly retrieve and emphasize the key RoIs that are crucial to the task improvements.

	WMT14 (EN-DE)		WMT14 (EN-FR)	
	EN→DE	EN←DE	EN→FR	EN←FR
+ Auto RoI	29.03	31.96	41.82	36.76
+ Gold RoI	**29.51**	**32.23**	**42.03**	**36.98**
Δ	-0.48	-0.27	-0.31	-0.22
	ParaNMT		QUORA	
	SRC→TGT	SRC←TGT	SRC→TGT	SRC←TGT
+ Auto RoI	31.53	30.60	38.66	37.58
+ Gold RoI	**31.86**	**30.85**	**39.02**	**38.11**
Δ	-0.33	-0.25	-0.36	-0.53

Table 5. Results (BLEU) of dual learning with automatically learned and gold RoI matching respectively.

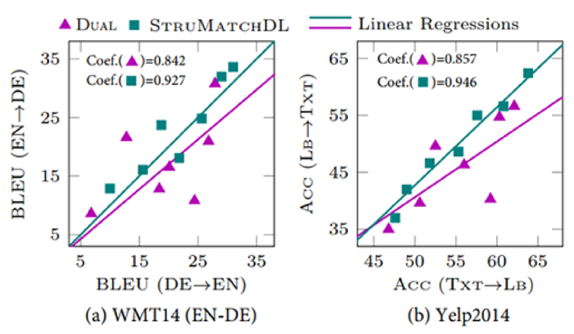

Figure 7. Performance correlation between two coupled tasks. 'Coef.' indicates Pearson correlation coefficient.

Evaluating Generated Text. ★ Second, our method strengthens the duality between two dual tasks by correctly aligning the RoIs.

	ParaNMT			MsCoCo		
	Gram.	Corr.	Cont.	Gram.	Corr.	Cont.
Human	4.86	4.92	3.78	4.82	4.15	4.37
Baseline	1.58	2.20	1.04	0.78	1.23	0.98
Dual	2.24	2.55	1.46	1.80	2.38	1.25
StruMatchDL	**3.78***	**3.67***	2.51	**3.46***	**3.27***	2.74
-SyRec	2.89	3.21	**2.90***	2.75	2.89	**2.96***

Table 7. Human evaluation results. Grammaticality (Gram.), correctness (Corr.), and content richness (Cont.) are rated on Likert 5-scale. $*$ indicates significantly better over the variant (p<0.03).

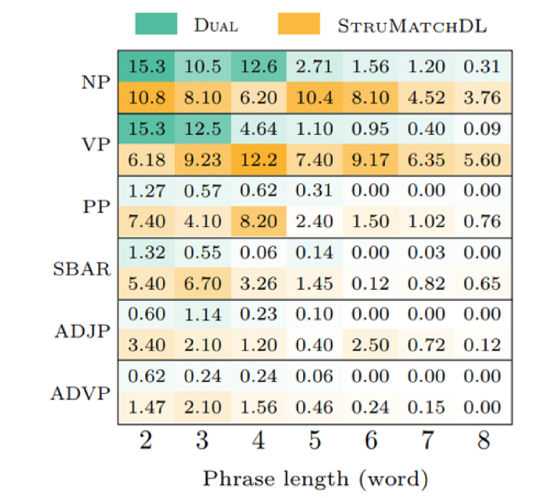

Figure 8. Distribution (frequency, %) over different constituency length of phrases in the generated sentences.

Evaluating Extendibility. ★ Third, the success of structure matching can be extended to non-text↔non-text dual learning, besides NLP tasks.

	CIFAR-10			CIFAR-100		
	Img→Lb	Img←Lb		Img→Lb	Img←Lb	
	ACC	IS↑	FID↓	ACC	IS↑	FID↓
M1	93.05	8.62	13.53	72.60	9.34	19.63
M3	93.68	9.83	9.80	73.85	13.64	15.72
M4	94.74	10.64	7.38	74.63	14.65	13.42
Δ	+1.06	+0.81	-2.42	+0.78	+1.01	-2.30

Table 10. Image↔Label experiment (IMG→LB: image classification, IMG←LB: conditioned image generation) on CIFAR-10 and CIFAR-100 datasets.

	CelebA-HQ		AFHQ	
	Img$_A$ →Img$_B$	Img$_A$ ←Img$_B$	Img$_A$ →Img$_B$	Img$_A$ ←Img$_B$
M1	26.7	32.7	32.4	40.8
M3	20.0	24.6	26.2	29.6
M4	17.5	20.3	22.0	25.7
Δ	-2.5	-4.3	-4.2	-3.9

Table 11. Image↔Image experiment (image-image translation) on CelebA-HQ and AFHQ datasets. Metrics: FID↓.

Insights into Key Influencers. ★ Fourth, the richer the structural information for the alignment, the better the improvements our method presents.

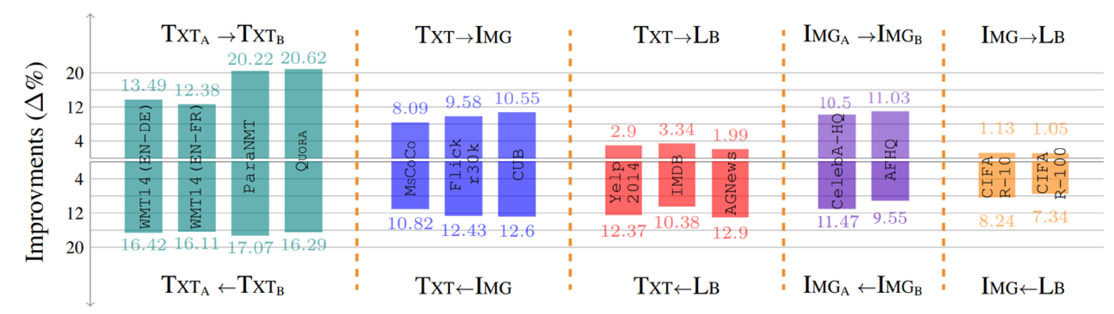

Figure 6: Relative task performance growth rates (Δ%) after taking the structure matching for dual learning.