Automatic Paraphrasing via Sentence Reconstruction and Round-trip Translation

IJCAI 2021 MONTREAL

Zilu Guo[1], Zhongqiang Huang[2], Kenny Q. Zhu[1], Guandan Chen[2], Kaibo Zhang[2], Boxing Chen[2] and Fei Huang[2]
[1] Shanghai Jiao Tong University
[2] Alibaba Damo Academy

Abstract

Paraphrase generation plays key roles in NLP tasks such as question answering, machine translation, and information retrieval. In this paper, we propose a novel framework for paraphrase generation. It simultaneously decodes the output sentence using a pretrained wordset-to-sequence model and a round-trip translation model. We evaluate this framework on Quora, WikiAnswers, MSCOCO and Twitter, and show its advantage over previous state-of-the-art unsupervised methods and distantly-supervised methods by significant margins on all datasets. For Quora and WikiAnswers, our framework even performs better than some strongly supervised methods with domain adaptation. Further, we show that the generated paraphrases can be used to augment the training data for machine translation to achieve substantial improvements.

Introduction

Paraphrase: a pair of sentences with similar meaning, but different wording.

Two kinds of underlying semantics:

1. Word set
2. The translation in another language

word set: **(man, sit, bike, bench)**
A *man* is *sitting* on a *bench* next to a *bike*
A *man* is *sitting* next to a *bench* next to a *bicycle*
A *man sits* on a *bench* by a *bike*
Man sitting on a *bench* near a personal *bicycle*
A *man* is *sitting* on a *bench* with a *bike*

Table 1. Paraphrases formed from a word set.

How to generate parpahrase:

Step 1. Generate a word set from the input
Step 2. Translate the input into another language
Step 3. Generate paraphrase through the word set and the translation with a hybrid decoder

Data Augmentation For NMT (English -- X)

Step 1. Extract English sentences from the training pairs
Step 2. Generate paraphrase For English sentences
Step 3. Combine the paraphrase and the X language from the original training pair to get new training pairs

Approach

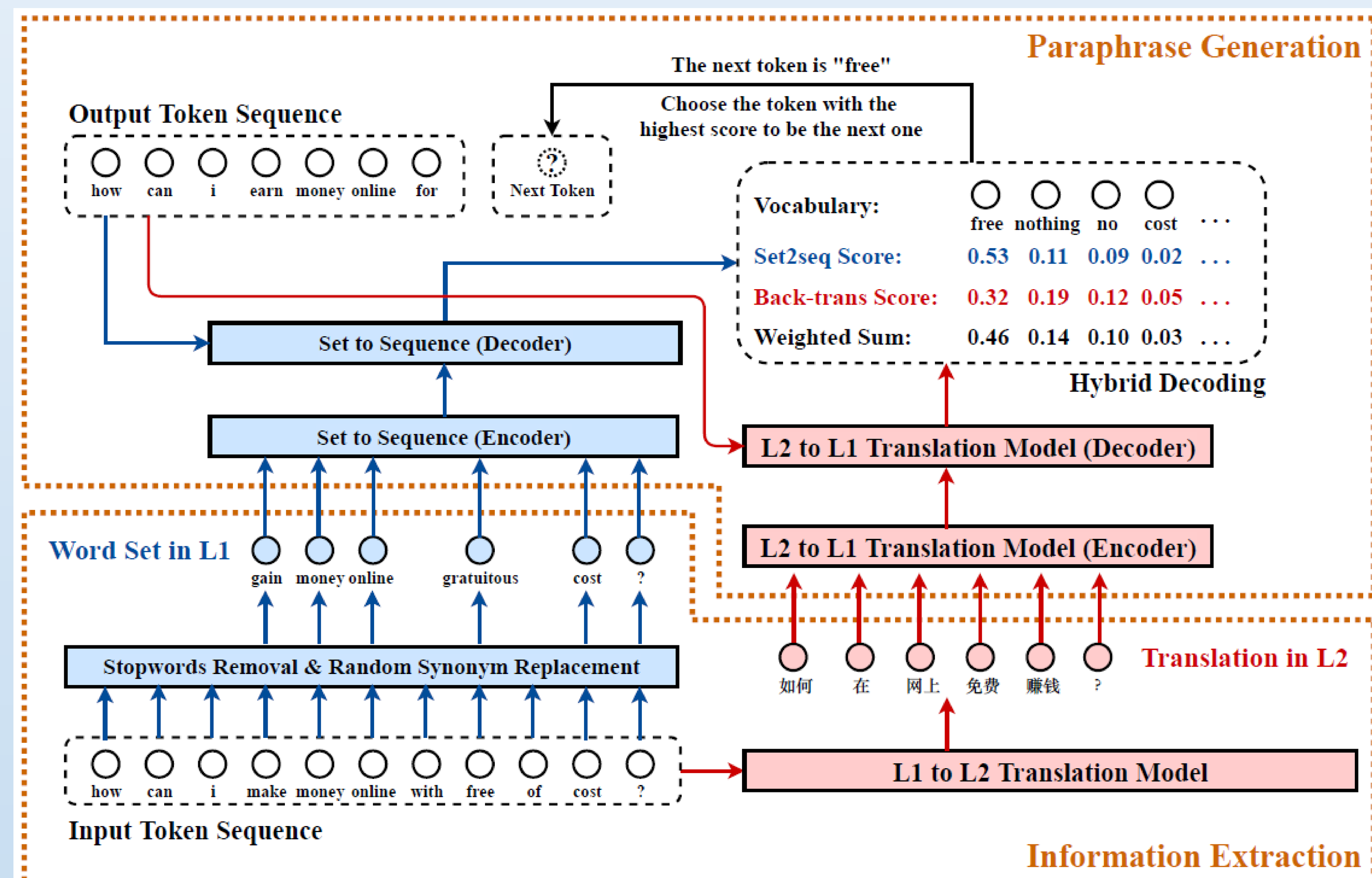

Figure 1. Our Paraphrasing Framework

Algorithm 1 Paraphrasing Framework

Input: Original sentence $X = [x_1, x_2, \cdots]$;
Output: Paraphrase $Y = [y_1, y_2, \cdots]$;

1: Reduce X to a set of keywords KWS by removing stopwords;
2: Obtain WS from KWS by random replacement with synonyms;
3: Translate X into Language L_2: $Z = [z_1, z_2, \cdots]$;
4: Encode WS with set2seq to hidden state H_{ws};
5: Encode Z with L_2-L_1 translation model to hidden state H_{bt};
6: Initialize: Y = [], y_0 = BOS, t = 0;
7: **while** y_t != EOS and t <length-limit **do**
8: $t = t + 1$;
9: Calculate y_t with Eqn. 3;
10: Y.append(y_t);
11: **end while**
12: **return** Y;

Figure 2. Detail For our framework

Results

Set2seq model:
Trained with in-domain & non-parallel data
Trained on a single GTX-2080 GPU for 3 hours

Round-trip Translation model:
Trained with WMT17 parallel zh-en dataset
Trained on two GTX-2080 GPUs for 3 days

Set2seq-common model:
Trained with WMT17 English monolingual data
Trained on a single GTX-2080 GPU for 1.5 days

	Model	Quora iBLEU	Quora BLEU	Quora R-1	Quora R-2	WikiAnswers iBLEU	WikiAnswers BLEU	WikiAnswers R-1	WikiAnswers R-2
Supervised	DNPG (SOTA)	18.01	25.03	63.73	37.75	34.15	41.64	57.32	25.88
Supervised + Domain-Adapted	Pointer-generator	5.04	6.96	41.89	12.77	21.87	27.94	53.99	20.85
	Transformer+Copy	6.17	8.15	44.89	14.79	23.25	29.22	53.33	21.02
	MTL+Copy	7.22	9.83	47.08	19.03	21.87	30.78	54.10	21.08
	DNPG	10.39	16.98	56.01	28.61	25.60	35.12	56.17	23.65
Unsupervised	CGMH	9.94	15.73	48.73	26.12	20.05	26.45	43.31	16.53
	UPSA	12.02	18.18	56.51	30.69	24.84	32.39	54.12	21.45
Distantly-Supervised	Liu *et al.* [2020]	9.90	15.03	52.65	23.18	-	-	-	-
	ParaNMT(round-trip translation)	10.69	15.75	52.28	25.12	14.94	20.01	30.55	10.23
	ParaBank	9.92	14.71	50.03	23.80	13.14	17.56	28.97	9.34
	set2seq (ours)	13.54	20.85	58.27	32.59	25.98	33.41	55.95	23.08
	set2seq-common+RTT (ours)	12.60	18.85	57.13	31.19	25.04	33.43	55.81	23.12
	set2seq+RTT (ours)	**14.66**	**22.53**	**59.98**	**34.09**	**28.27**	**37.42**	**56.71**	**24.94**

	Model	MSCOCO iBLEU	MSCOCO BLEU	MSCOCO Rouge1	MSCOCO Rouge2	Twitter iBLEU	Twitter BLEU	Twitter Rouge1	Twitter Rouge2
Unsupervised	CGMH	7.84	11.45	32.19	8.67	4.18	5.32	19.96	5.44
	UPSA	9.26	14.16	37.18	11.21	4.93	6.87	28.34	8.53
Distantly-Supervised	Liu *et al.* [2020]	6.67	9.86	22.14	6.21	-	-	-	-
	ParaNMT(round-trip translation)	7.39	10.71	30.74	8.68	7.57	10.79	35.38	14.74
	ParaBank	6.45	9.48	29.22	8.35	6.50	9.71	34.56	13.92
	set2seq (ours)	**11.54**	17.61	39.87	13.67	5.72	7.48	31.65	10.89
	set2seq-common+RTT (ours)	9.07	13.44	35.90	11.05	9.73	**14.30**	**39.23**	**18.82**
	set2seq+RTT (ours)	11.39	**17.93**	**40.28**	**14.04**	**9.95**	13.97	38.96	18.32

Table 2. Compared with baseline methods

Model Variants	iBLEU	$BLEU_{ref}$	$BLEU_{src}$
set2seq+RTT	**14.66**	22.53	**56.17**
⊖ excluding stopwords ⊕ retaining high-IDF	13.46	22.15	64.75
⊖ random replacement	13.78	**23.92**	77.47
⊕ position encoding	14.07	23.26	68.60

Table 3. Ablation Study

Method	Accuracy Score	Accuracy Agreement	Fluency Score	Fluency Agreement
CGMH	3.15	0.55	3.42	0.50
UPSA	3.49	0.54	3.51	0.55
DNPG(Adapted)	3.32	0.48	3.62	0.54
RTT	3.37	0.59	**4.18**	0.58
set2seq+RTT(ours)	**3.78**	0.57	4.13	0.55

Table 4. Human evaluation

	Size	Orig. Pairs	Augmented
De-En	150k	12.89	15.06
	300k	15.67	17.20
Zh-En	150k	10.21	11.99
	300k	12.10	14.07
Ru-En	150k	16.88	18.55
	300k	19.30	21.09

Table 5. Results For NMT data augmentation