Paraphrase Generation with Latent Bag of Words

Yao Fu[1], Yansong Feng[2] and John P. Cunningham[1]

[1] Columbia University
[2] Peking University

Task Definition

How do I speak English fluently

↓

How can I improve my spoken English

- Paraphrase Generation: to change the expression of a sentence while conveying the same meaning

Step-by-step Generation

	How	do	I	speak	English	fluently
		↓		↓	↓	↓
Word neighbors		can may could		speaking spoken oral	language	fluency well improve

Generation based on neighbor words →

How may I speak English fluently
How can I improve my spoken English
How could I improve my English fluency

Traditional planning and realization:

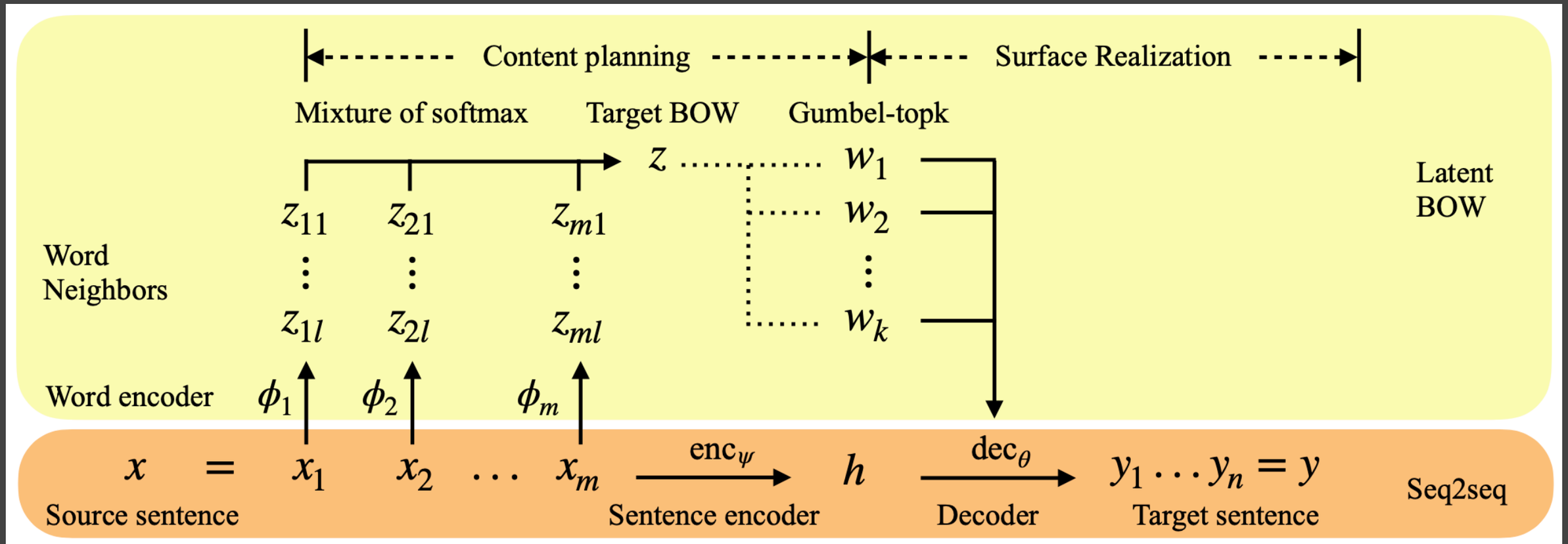
- neighbor words from WordNet
- sample from neighbor words
- rule-based realization

→

This work:

- infer neighbors as latent variables
- differentiable subset sampling
- end-to-end generation

The Latent BOW Model



- Infer word neighbors for each input words
- Ground inferred neighbors w. target BOW
- Differentiable subset sampling from inferred word neighbors w. Gumbel-topk reparameterization
- Generate sentence based on sampled neighbors and the input sentence

Performance Gain w. Latent BOW

Quora

Model	B-1	B-2	B-3	B-4	R-1	R-2	R-L
Seq2seq[39]	54.62	40.41	31.25	24.97	57.27	33.04	54.62
Residual Seq2seq-Attn [39]	54.59	40.49	31.25	24.89	57.10	32.86	54.61
β-VAE, $\beta = 10^{-3}$[16]	43.02	28.60	20.98	16.29	41.81	21.17	40.09
β-VAE, $\beta = 10^{-4}$[16]	47.86	33.21	24.96	19.73	47.62	25.49	45.46
BOW-Hard (lower bound)	33.40	21.18	14.43	10.36	36.08	16.23	33.77
LBOW-Topk (ours)	**55.79**	**42.03**	**32.71**	**26.17**	**58.79**	**34.57**	**56.43**
LBOW-Gumbel (ours)	55.75	41.96	32.66	26.14	58.60	34.47	56.23
RbM-SL[25]	-	43.54	-	-	64.39	38.11	-
RbM-IRL[25]	-	43.09	-	-	64.02	37.72	-
Cheating BOW (upper bound)	72.96	61.78	54.40	49.47	72.15	52.61	68.53

MSCOCO

Model	B-1	B-2	B-3	B-4	R-1	R-2	R-L
Seq2seq[39]	69.61	47.14	31.64	21.65	40.11	14.31	36.28
Residual Seq2seq-Attn [39]	71.24	49.65	34.04	23.66	41.07	15.26	37.35
β-VAE, $\beta = 10^{-3}$[16]	68.81	45.82	30.56	20.99	39.63	13.86	35.81
β-VAE, $\beta = 10^{-4}$[16]	70.04	47.59	32.29	22.54	40.72	14.75	36.75
BOW-Hard (lower bound)	48.14	28.35	16.25	9.28	31.66	8.30	27.37
LBOW-Topk (ours)	**72.60**	**51.14**	**35.66**	**25.27**	42.08	**16.13**	**38.16**
LBOW-Gumbel (ours)	72.37	50.81	35.32	24.98	**42.12**	16.05	38.13
Cheating BOW (upper bound)	80.87	75.09	62.24	52.64	49.95	23.94	43.77

Clearly Interpretable Generation Steps

Quora

Input	why	do	people	ask	questions	on	quora	instead	of	googling	it		
Neighbor				*post*	*quora*		*quora*			*google*			
				answer	*questions*		*questions*			*search*			
BOW sample	*ask, quora, people, questions, google, googling, easily, googled, search, answer*												
Output	why do people ask questions on quora that can be easily found on a google search ?												
Input	how	do	i	talk	english	fluently	?						
Neighbor				*speak*	*english*	*fluently*							
				better	*improve*	*confidence*							
BOW sample	*english, speak, improve, fluently, talk, spoken, better, best, confidence*												
Output	how can i improve my english speaking ?												

MSCOCO

Input	A	tennis	player	is	walking	while	holding	his	racket				
Neighbor		*court*	*holding*		*walks*		*carrying*		*court*				
		racket	*man*		*across*		*holds*		*racquet*				
BOW sample	*holding, man, tennis, walking, racket, court, player, racquet, male, woman, walks*												
Output	A man holding a tennis racquet on a tennis court												
Input	A	big	airplane	flying	in	the	blue	sky					
Neighbor		*large*	*airplane*	*sky*			*blue*	*clear*					
		large	*jet*	*airplane*			*clear*	*flying*					
BOW sample	*blue, airplane, flying, large, plane, sky, clear, air, flies, jet*												
Output	A large jetliner flying through a blue sky												

word morphology | synonym | entailment | metonymy

Conclusion

- Performance gain with LBOW
- Unsupervised learning of word neighbors
- Interpretable step-by-step generation
- Code available at https://github.com/FranxYao/dgm_latent_bow