

Image Description Dataset for Language Learners (LREC2022)

Kento Tanaka, Taichi Nishimura, Hiroaki Nanjo, Keisuke Shirai,
Hirota Kameko and Masatake Dantsuji
Kyoto University, Kyoto Tachibana University

Summary

Image Description Task

➢ A learner is asked to produce a sentence that describes an image



Grammatical Error Correction (GEC)

➢ Task to correct grammatical errors in a sentence

➢ GEC system cannot correct specific errors in image description tasks (i.e. semantic or pragmatic errors)



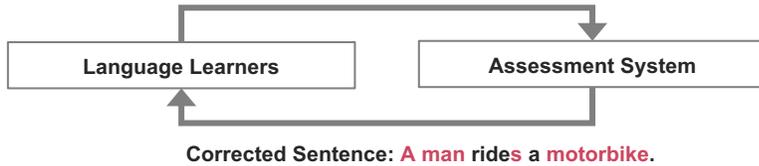
Contributions

- 1) Dataset Construction
 - Construct a new dataset for automatic assessment of image description
- 2) Error Correction
 - Propose a novel task of automatic error correction for image description

Given Image:



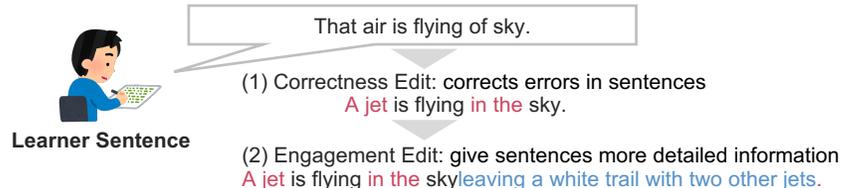
Learner Sentence: Men ride a bicycle.



Dataset Construction

Collection Learner Sentences

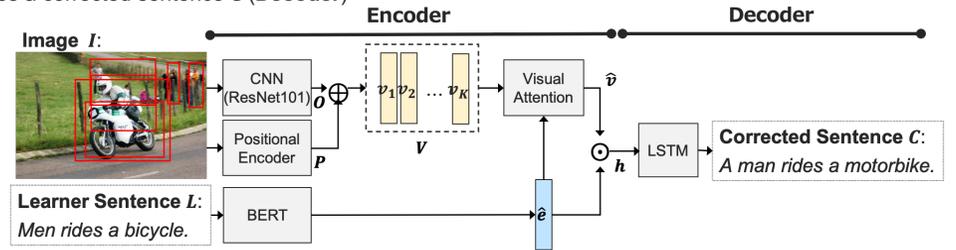
- Conduct the exercises with Japanese high school students who learns English
- As a result, collected 651 sentences for 120 images



Error Correction

Processes

- 1) Converts objects in an image I into object-wise representation (**Image Encoder**)
- 2) Converts a learner sentence L into a sentence representation (**Sentence Encoder**)
- 3) Calculates an attention vector \hat{v} (**Visual Attention**)
- 4) Generates a corrected sentence C (**Decoder**)



Experiments

Data Limitation

- Constructed LLID size is limitation
 - Generate synthetic dataset on the corrected sentence to pre-train the model

Synthetic Error Generation

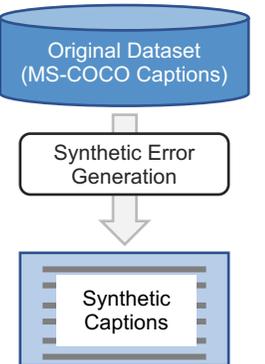
- Follow the existing approach [Awasthi+, IJCNLP19] and apply it on MS-COCO captions

Dataset

- Pre-train: MS-COCO Captions + Synthetic Captions
- Fine-tune: Learner Sentence + Correctness on LLID

Model

- L-C: Generate a corrected sentence (C) from a learner sentence (L) w/o an image
- LI-C: Generate a corrected sentence (C) from a learner sentence (L) w/ an image (I)



Quantitative Evaluation

Model	Image Features	ERRANT			GLEU
		Prec.	Rec.	$F_{0.5}$	
GECToR	-	0.367	0.116	0.256	0.268
LI-C	Global	0.147	0.168	0.151	0.263
LI-C	Bottom-up	0.165	0.190	0.170	0.284

Ablation Study

Model	Image Features	Synthetic Dataset	ERRANT			GLEU
			Prec.	Rec.	$F_{0.5}$	
L-C	-	✓	0.105	0.126	0.108	0.226
LI-C	Bottom-up		0.137	0.183	0.144	0.176
LI-C	Bottom-up	✓	0.165	0.190	0.170	0.284

Qualitative Evaluation

(1)		Learner Sentence	That air is flying of sky.
		Ground Truth	A jet is flying in the sky.
		GECToR	That air is flying of sky.
(2)		LI-C (Bottom-up)	A fighter is flying in the sky.
		Learner Sentence	A old man is having wain grass.
		Ground Truth	An old man is holding a wine glass.
		GECToR	An old man is having wain grass.
		LI-C (Bottom-up)	A man is having a conversation.