The AIP-Tohoku System at the BEA-2019 Shared Task

Hiroki Asano[12]*, Masato Mita[21], Tomoya Mizumoto[21]†, Jun Suzuki[12]

[1]Tohoku University, [2]RIKEN Center for Advanced Intelligence Project (AIP), *Yahoo Japan Corporation, †Future Corporation
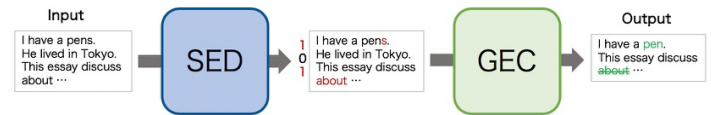
▊ Summary

- This is the first study that has combined GEC with sentence-level error detection (SED)
- Our result demonstrates SED improve the precision of GEC
- Our system is ranked 9th in Track1 and 2nd in Track2

System Architecture

We input grammatically incorrect sentences predicted by the SED model into our GEC model
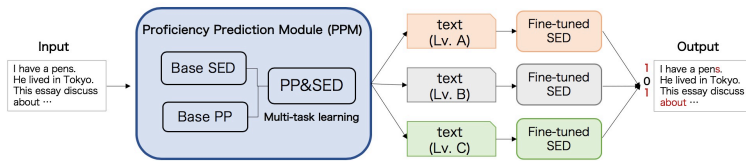

🔍 Key Technique: Sentence-level Error Detection (SED)

▊ Motivation

Reduce FP by passing only sentences that contain errors to the GEC model using SED

▊ Architecture

	Prec.	Rec.	F
Base SED	88.5	79.8	83.9
Proposed SED	91.3	95.6	93.4

fine-tuned (+9.5 F point)

Base SED
- Performs sentence-level binary classification of sentences that need editing

Proficiency Prediction Module (PPM)
- Base PP predicts the leaners proficiency
- Employed a multi-task learning approach in which PP model and SED model simultaneously

Fine-tuned SED
- SED model is fine-tuned for each level of proficiency (Lv. A, Lv. B, Lv. C)

📊 Experimental Configurations

▊ Model

GEC Model
- Transformer-based Model

SED Model
- BERT-based Model

Error Generation Model (GenData)
- Following the system by Edunov et al. (2018)

▊ Dataset

Model	Track1	Track2
GEC	• Official data (564K)	• Official data (564K) • EFCAMDAT [Geertzen et al+2013] + Non-public Lang-8 (7.7M)
GenData	• Simple Wikipedia + Essay scoring data sets (i.e, ICLE [Granger+2009], ICNALE[Ishikawa], ASAP, TOEFL 11[Blanchard+2013]) (1.4M)	
SED	• Official data (564K)	

📊 Results

▊ Main Leaderboard

Model	Prec.	Rec.	F0.5	Rank
Track1	68.62	42.16	60.97	9th
Track2	70.60	51.03	65.57	2nd

▊ Ablation Test

Model	Prec.	Rec.	F0.5
GEC	61.97	42.11	56.63
+GenData	64.57	46.40	59.88
+SED	68.62	42.16	60.97

+ 4.05 point