

The LAIX Systems in the BEA-2019 GEC Shared Task

Ruobing Li Chuan Wang Yefei Zha Yonghong Yu Shiman Guo Qiang Wang Yang Liu Hui Lin
LAIX Inc.

Abstract

In this paper, we describe two systems we developed for the three tracks we have participated in the BEA-2019 GEC Shared Task. We investigate competitive classification models with bi-directional recurrent neural networks (Bi-RNN) and neural machine translation (NMT) models. For different tracks, we use ensemble systems to selectively combine the NMT models, the classification models, and some rules, and demonstrate that an ensemble solution can effectively improve GEC performance over single systems. Our GEC systems ranked the first in the Unrestricted Track, and the third in both the Restricted Track and the Low Resource Track.

Data

Track	FCE	Lang-8	NUCLE	W&I+LOCNESS	Common Crawl	Wikied	Wiki dumps
Restricted Track	Yes	Yes	Yes	Yes	Yes	-	-
Unrestricted Track	Yes	Yes	Yes	Yes	Yes	-	-
Low Resource Track	-	-	-	-	Yes	Yes	Yes

Restricted Track

- CNN-based ensemble translation systems

Ensemble index	Combination	Precision	Recall	$F_{0.5}$
1	5012,5102,7011,7205	0.5076	0.2195	0.4021
2	5001,5002,5003,5004	0.5003	0.1951	0.3811
3	5005,5012,5102,7205	0.5156	0.2150	0.4029
4	5005,5012,7011,7205	0.5152	0.2159	0.4034

- Transformer-based translation ensemble systems

Model index	Error weight	Copy number of W&I trainset	GPU number	Precision	Recall	$F_{0.5}$
1	3	10	2	0.4585	0.3525	0.4325
2	3	10	4	0.4602	0.3514	0.4333
3	3	8	2	0.4592	0.3575	0.4345
4	3	8	4	0.4641	0.3548	0.4372
5	3	15	2	0.4494	0.3479	0.4247
6	3	15	4	0.4648	0.3467	0.4352
7	2	10	2	0.4715	0.3303	0.4343
8	2	10	4	0.4868	0.3412	0.4485

Conflict Solver

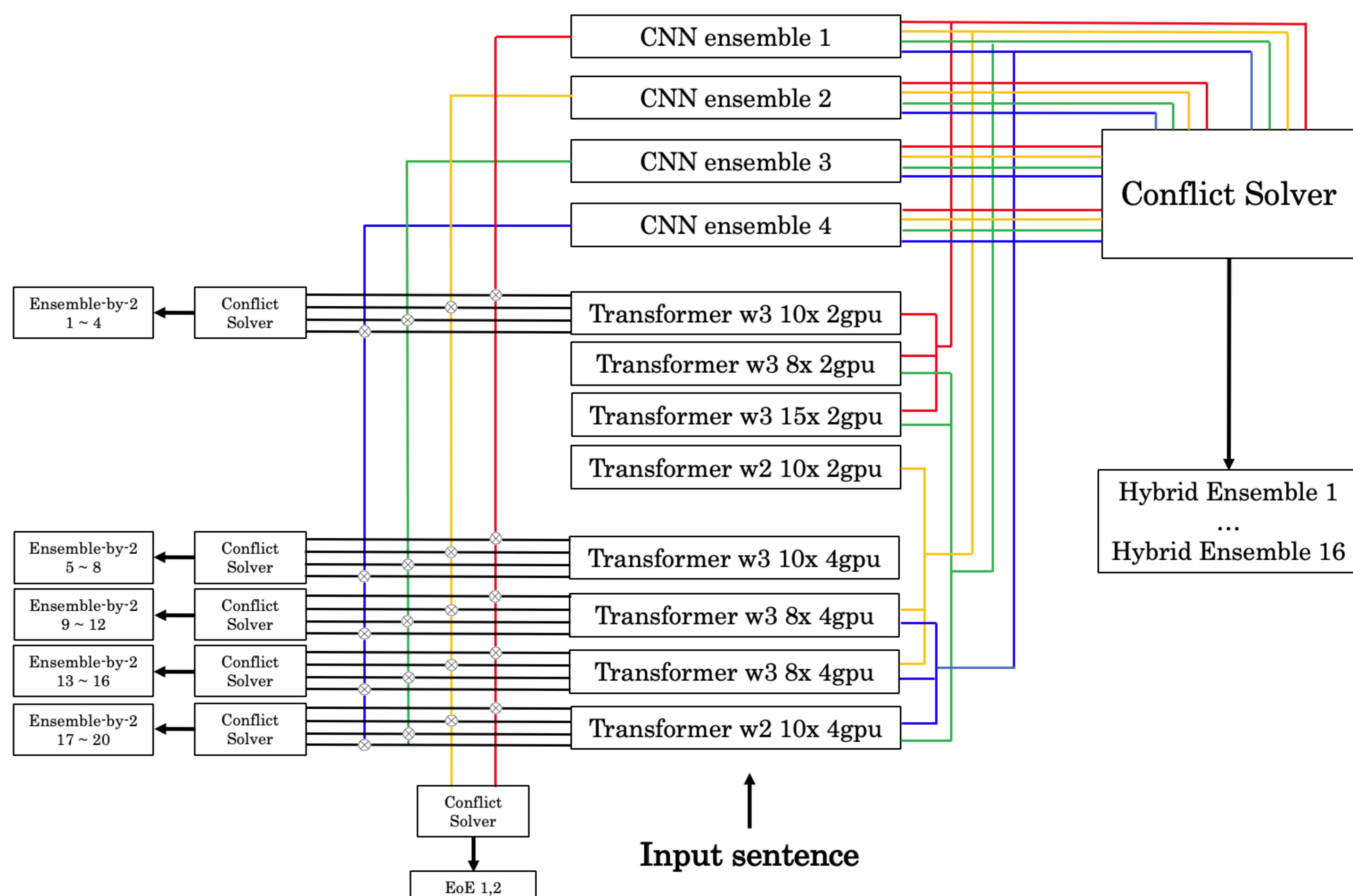
- When combining outcomes from different systems, we treat the precision in a confidence table as the confidence. Each correction has its confidence obtained by looking up the precision of the corresponding type of the correction in the table. If two conflicting corrections are the same, we merge them and add α to the confidence of the correction; otherwise, the correction with a lower confidence will be discarded. In this way, we merge conflicting errors one by one.
- After combining outcomes, if the confidence of a correction is lower than β , the correction is discarded.
- γ is used to distinguish when it is more important to focus on the precision or $F_{0.5}$ of a correction. When we move to the final ensemble with confidence tables of existing systems, if the confidence is larger than γ , we select the correction proposed by the system that has the best $F_{0.5}$ on the type of this correction. Otherwise, the correction by a system with the best precision is selected.

Parameters for all the conflict solvers

Ensemble method	α	β	γ
Ensemble-by-2	0.2	0.4	-
EoE	0.15	0.8	-
Hybrid ensemble	0.15	0.62	-
Final ensemble	0.0	0.5	0.52

Restricted and Unrestricted Track

- The architecture of the ensemble translation system in Restricted and Unrestricted Tracks



- Result for Restricted and Unrestricted Tracks

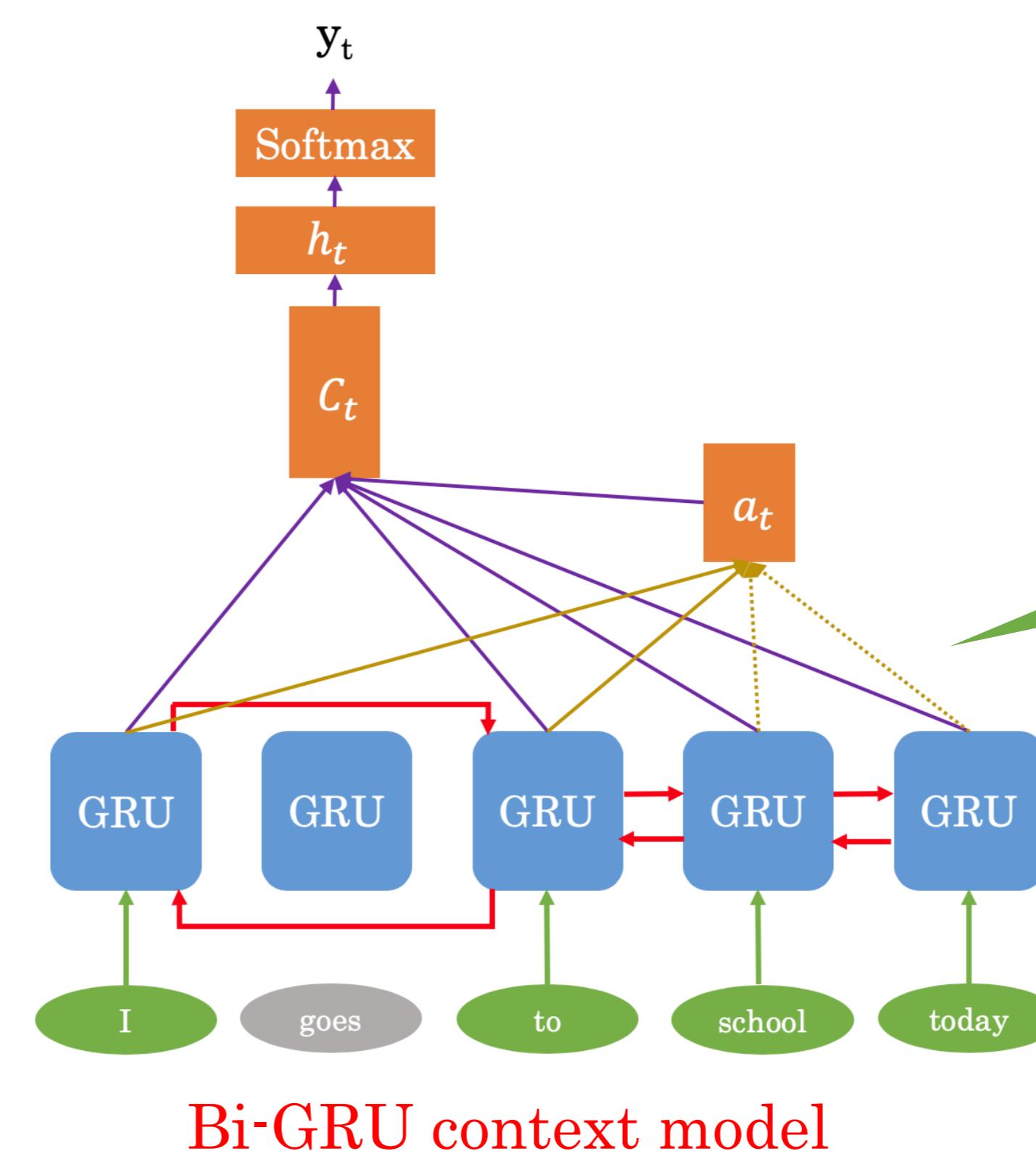
Step	Precision	Recall	$F_{0.5}$
Best CNN-based ensemble model	0.5152	0.2159	0.4034
Best Transformer-based translation model	0.4868	0.3412	0.4485
Best ensemble-by-2	0.5281	0.3434	0.4768
Best hybrid ensemble	0.5885	0.3278	0.5078
+ Combine best performance	0.6283	0.3269	0.5305
+ Second pass	0.6272	0.3412	0.5372
Submission system (+ Post-processing, Dev set)	0.6243	0.3457	0.5376
Submission system (Test set)	0.7317	0.4950	0.6678

Low Resource Track

- Classification models

✓ Bi-GRU context model
Subject-verb agreement, Article, Plural or singular noun, Verb form, Preposition substitution, Missing comma and Period comma

✓ Pointer context model
Word form



- Result for Low Resource Track

Model	Precision	Recall	$F_{0.5}$
Rule	0.4497	0.0216	0.0905
Spelling	0.3188	0.0363	0.1248
Article	0.4367	0.0134	0.0597
Missing comma	0.4729	0.0503	0.1763
Period comma substitution	0.4561	0.0070	0.0328
Plural or singular noun	0.3203	0.0121	0.0524
Preposition substitution	0.3713	0.0101	0.0454
Subject-verb agreement	0.3981	0.0115	0.0517
Verb form	0.4135	0.0074	0.0344
Word form	0.4506	0.0294	0.1164
NMT	0.1279	0.1480	0.1315
Submission system (Dev set)	0.4970	0.1686	0.3577
Submission system (Test set)	0.6201	0.3125	0.5181

Ensemble method	α	β	γ
Ensemble for all	0.15	0.3	-
Final ensemble	0.0	0.25	0.3

Low Resource Track

