

Selecting, Planning, and Rewriting: A Modular Approach for Data to Document Generation and Translation

Lesly Miculicich, Marc Marone, Hany Hassan

We present our system submissions to all 6 tracks of the WNGT 2019 shared task on Document-Level Generation and Translation.

Translation

For both the Translation and Translation + Generation Tracks, we focus on adapting the WMT19 Microsoft model of Xia et al.

	EN – DE	DE – EN
WMT19 Fairseq	36.26	42.91
WMT19 Microsoft	39.99	48.71
+ Fine-tuning	47.90 (+7.9)	57.99 (+9.3)
+ BT [†]	-	59.94 (+1.9)
+ Ensemble [†]	48.09 (+0.2)	-
Naver Labs [‡]	48.02	58.22

In early experiments, we found that adding the structured data did not seem to help much. As a result, we used the pure MT models for the MT + NLG Tasks as well.

Data + EN -> DE	RG	CS (P/R)	CO
WMT19 Fairseq	81.64	77.67 / 75.82	60.83
Microsoft-FT	80.98	76.88 / 84.57	67.84
Naver Labs	80.65	75.10 / 88.72	69.17

Data + DE -> EN	RG	CS (P/R)	CO
WMT19 Fairseq	93.53	83.33 / 84.22	70.47
Microsoft-FT	95.77	92.49 / 91.62	84.70
Naver Labs	95.63	91.71 / 92.69	85.05

Examples

GPT

External Knowledge

- The Milwaukee Bucks defeated the New York Knicks at **Madison Square Garden**
- The Cleveland Cavaliers defeated the Boston Celtics at the **TD Garden in Boston**
- The Toronto Raptors defeated the Philadelphia 76ers, 122 - 95, at **Air Canada Centre**

Multi Record Synthesis

- Milwaukee also got a nice game from Giannis Antetokounmpo , who finished with **16 points (6 - 9 FG , 1 - 1 3Pt , 3 - 6 FT)**, **12 rebounds** and a steal.
- Chris Paul had a great all-around game, as he dished out **20 assists** and scored **17 points** in the win. DeAndre Jordan scored **19 points**, while Marreese Speights finished with **17 points** and **12 rebounds** in **25 minutes** off the bench.

Negative Examples

- The Los Angeles Clippers defeated the New Orleans Pelicans 133 – 105
- The Milwaukee Bucks (**10**) defeated the New York Knicks (1 - 0) **97 - 122** on Friday
- The Raptors were able to prevail despite **being outshot, out rebounded and out – rebounded**

Sampling

Top-90

Terrence Ross had to carry the load in the absence of DeRozan, as Ross finished with a game - high **22 points** on **8-of-11** shooting, including **3-of-5** from long range. Lowry added **24 points, eight assists and four rebounds** in **32 minutes**.

Top-90

In overtime, the Raptors took a 13 - point lead into the locker room thanks to a clutch 30 - foot field goal from Terrence Ross with eight seconds remaining in regulation. Ross finished with a game - high **22 points** on **8-of-11** shooting, including a **3-of-5** shooting from long range.

Top-50

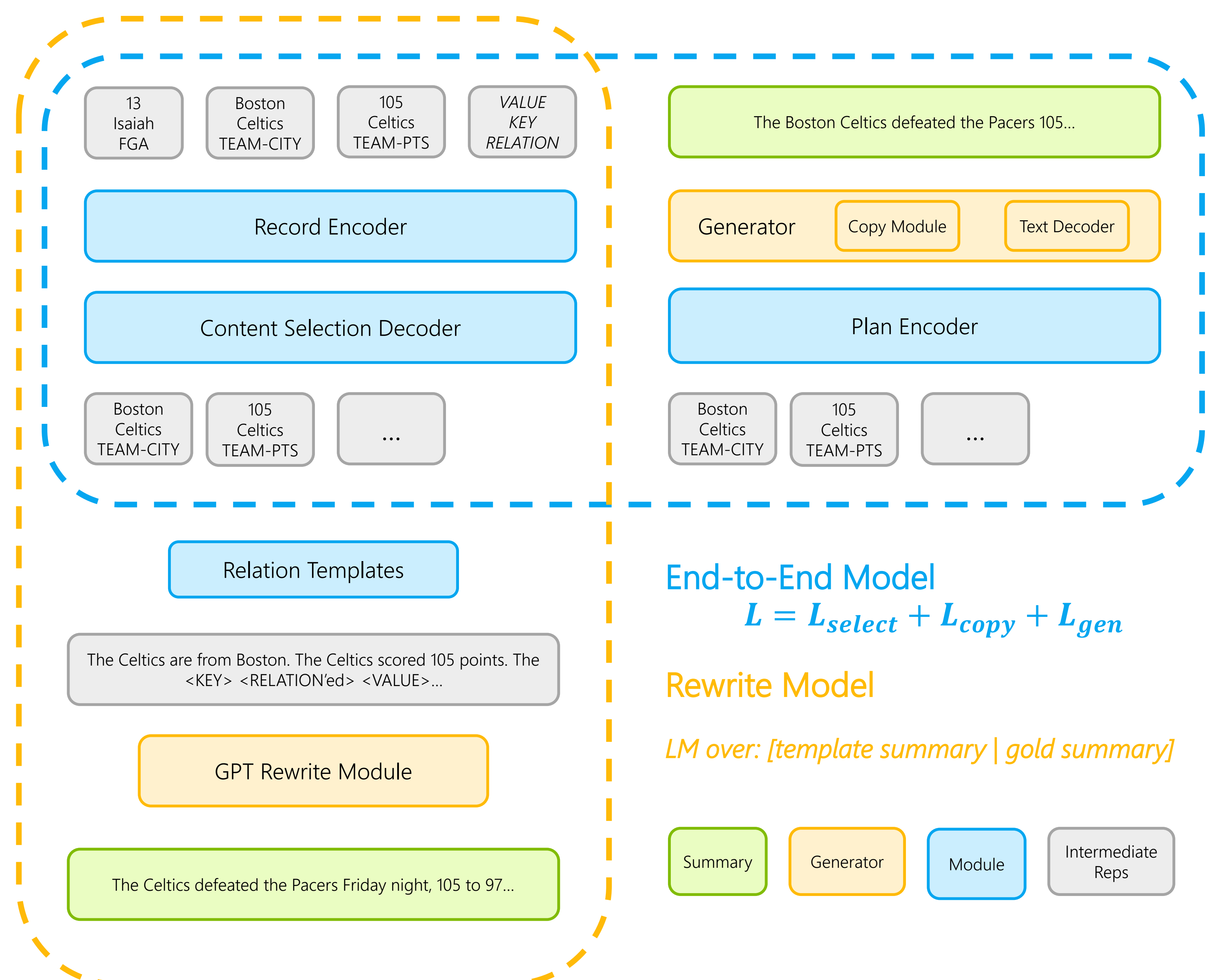
The Raptors were led by Terrence Ross, who posted a game - high **22 points** on **8-of-11** shooting, including **3-of-5** from long range. Kyle Lowry followed up with **24 points, eight assists and four rebounds, in 32 minutes**.

Gold

Kyle Lowry carried the load for the Raptors, accumulating **24 points, four rebounds and eight assists**. Terrence Ross was great off the bench, scoring **22 points** on **8-of-11** shooting.

Generation

We build on the Content Selection and Planning model of Puduppully et al. We add self-attention in the record encoders, a shared value and summary vocab, and two different methods for generating text from a content plan.



NLG	BLEU	RG	CS (P/R)	CO
E2E	15.03	93.38	32.40 / 58.02	18.54
GPT-50	15.17	94.35	33.91 / 53.82	19.30
GPT-90	13.03	88.70	32.84 / 50.58	17.36
Naver Labs	20.52	94.08	41.13 / 54.20	25.64

E2E

External Knowledge

- The Milwaukee Bucks defeated the New York Knicks at the **Staples Center in New Orleans**
- The Cleveland Cavaliers defeated the Boston Celtics at the **Wells Fargo Center in Boston**
- The Toronto Raptors defeated the visiting Philadelphia 76ers at **United Center**

Multi Record Synthesis

- The Bucks were led by Giannis Antetokounmpo , who finished with **16 points (6 - 9 FG, 1 - 1 3Pt, 3 - 6 FT) and 12 rebounds**.
- Chris Paul dished out a team-high **20 assists**, while DeAndre Jordan added **19 points** of his own. Marreese Speights came off the bench to score **17 points on 6-of-9 shooting...**

Negative Examples

- The **Los Angeles Clippers** defeated the **Los Angeles Pelicans** 133 – 105
 - The Milwaukee Bucks (0 - 1) defeated the **Los Angeles Knicks** (1 - 0) **97 - 122** on Saturday
- The Raptors shot **55 percent** from the field and **68 percent** from long range, while the 76ers shot **55 percent** from the field and **68 percent...**

Discussion

- Hard to find a balance between fluency and faithfulness to data
- Many errors can be traced to selection failures
- GPT further decouples selection and generation with some benefits, but at risk of hallucinations
- Some of the most glaring errors are logical or mathematical