

ELL 884

Deevyansh Khadria(2022EE31883)

April 2025

1 Why T5-base is Better than BART-large for Grammatical Error Correction

T5-base and BART-large are both sequence-to-sequence transformer models designed for text generation tasks. While BART-large has strong capabilities, T5-base offers certain advantages that make it particularly well-suited for grammatical error correction (GEC):

- **Unified Text-to-Text Framework:** T5 is trained under a consistent "text-to-text" paradigm, where every NLP task is framed as a text generation problem. This provides a natural advantage for GEC, which involves generating corrected sentences from incorrect ones. BART, on the other hand, was primarily trained for denoising autoencoding and may require more task-specific fine-tuning more.
- **Better Handling of Structured Outputs:** Due to its span-corruption objective and the structured nature of its pretraining, T5 tends to produce more grammatically coherent and syntactically complete outputs, which is crucial for GEC.

2 Best Hyperparameters for the Model:

- T5 base model and PEFT
- r=16
- lora alpha=16
- target modules=["q", "v"]
- lora dropout=0.05
- per device eval batch size=16
- learning rate=1e-4
- num train epochs=7,

- weight decay=0.01
- logging steps=8

3 BLEU Score

The BLEU (Bilingual Evaluation Understudy) score is a metric for evaluating the quality of machine-generated translations compared to human translations. It computes precision by comparing n-grams (up to a specified length, usually 4) between the candidate translation and reference translations. The score is calculated as:

$$\text{BLEU} = \text{BP} \times \exp \left(\sum_{n=1}^N w_n \cdot \log(p_n) \right)$$

Where p_n is the precision for n-grams, and BP is a brevity penalty to penalize short translations. The weight w_n for each n-gram is typically set to $\frac{1}{N}$.

NOTE: These are results for the 90-10 train validation split

Model	BLEU Score (%)	Exact Match Score (%)
T5-base	75.09	10.68
BART-large	70.09	5.67

Table 1: Validation BLEU and Exact Match Scores for T5-base and BART-large

4 GLEU Score

The GLEU (Google-BLEU) score is a metric designed to evaluate grammatical error correction and translation quality. Unlike BLEU, GLEU considers both precision and recall of n-grams between the candidate and references. It penalizes both missing and over-generated n-grams. The score is defined as:

$$\text{GLEU} = \min(\text{precision}, \text{recall})$$

Where precision and recall are computed over n-grams (typically up to 4-grams) in the candidate and reference sentences.