|  |  |
| --- | --- |
| Paper Title: | FrugalScore: Learning Cheaper, Lighter and Faster Evaluation Metrics for Automatic Text Generation |
| Paper DoI: | [10.18653/v1/2022.acl-long.93](http://dx.doi.org/10.18653/v1/2022.acl-long.93) |
| Research Question | Does FrugalScore perform at par or outperforms BERTScore and MoverScore? |
| Related Work | With the advent of transfer learning in NLP, new NLG evaluation metrics have come into practice such as the BERTScore (Zhang et al. 2019) and MoverScore (Zhao et al., 2019). However, these metrics require more time to compute and require access to significant computational resources due to the huge number of underlying models. Hence, are not easily accessible to everyone. |
| Experiment Design | **Approach:**  To address the problem seen in the aforementioned metric, the authors have proposed a light weight version of those metric called FrugalScore.  FrugalScore has three phases:   1. Phase 1: Create a synthetic dataset by sampling pairs of related sequences and annotating them with the expensive metrics to be learned. 2. Phase 2: They continue the pretraining of the miniature pretrained language model on the synthetic dataset build by Phase 1. Here the model learns the internal mapping of the expensive metric. 3. Phase 3: They then fine-tuned the miniature on human annotations which can boost performance of the model. |
| Result | FrugalScore has compact models with fewer parameters and hence runs faster compared to the original metrics. When working on summarization and translation, FrugalScore is on par with the original metrics. This newly trained metrics is freely available and easily accessible via Hugging Face to benefit the entire NLP community. |