Discriminative Adversarial Search for Abstractive Summarization Supplementary Material

Thomas Scialom [1 2] **Paul-Alexis Dray** [1] **Sylvain Lamprier** [2] **Benjamin Piwowarski** [2 3] **Jacopo Staiano** [1]

1. Implementation details

All models are implemented in PyText (Aly et al., 2018). For all our experiments we used a single RTX 2080 Ti GPU.

To train the discriminator, we used the Adam optimiser with the recommended parameters for BERT: learning rate of $3e^{-5}$, batch size of 4 and accumulated batch size of 32. We trained it for 5 epochs; each epoch took 100 minutes on 150k samples. During discriminator retraining, the generator is needed and thus additional memory is required: all else equal, we decreased the batch size to 2. The self-training process takes one epoch to converge, in about 500 minutes: 200 minutes for training the discriminator and 300 minutes to generate the summaries with the search procedure described in Algorithm 1.

References

Aly, A., Lakhotia, K., Zhao, S., Mohit, M., Oguz, B., Arora, A., Gupta, S., Dewan, C., Nelson-Lindall, S., and Shah, R. Pytext: A seamless path from nlp research to production. *arXiv preprint arXiv:1812.08729*, 2018.

[1]reciTAL, Paris, France [2]Sorbonne Université, CNRS, LIP6, F-75005 Paris, France [3]CNRS, France. Correspondence to: Thomas Scialom <thomas@recital.ai>.

Proceedings of the 37^{th} International Conference on Machine Learning, Online, PMLR 119, 2020.