

1 On the State of the Art of Evaluation in Neural Language Models

(Melis, Dyer, and Blunsom 2017)

Task: Word Level Language Modelling Datasets: PTB, WikiText-2

Optimizer: [Adam] $\beta_1 = 0$, $\beta_2 = 0.999$, $\epsilon = 10^{-9}$ **Batch size:** 64

- Learning rate multiplied by 0.1 whenever validation performance does not improve during 30 consecutive checkpoints (which are performed every 100 and 200 steps for PTB and WT2).

Task: Character Level Language Modelling

- Truncated backprop after 50 timesteps
- Checkpoints every 400 timesteps

Optimizer: [Adam] $\beta_2 = 0.99$, $\epsilon = 10^{-5}$ **Batch size:** 128

Melis, Gábor, Chris Dyer, and Phil Blunsom. 2017. “On the State of the Art of Evaluation in Neural Language Models.” *arXiv Preprint arXiv:1707.05589*.