

GPT-2 pretrained by size, train loss

