LLM 无标签微调

Semi-supervised Fine-tuning for Large Language Models

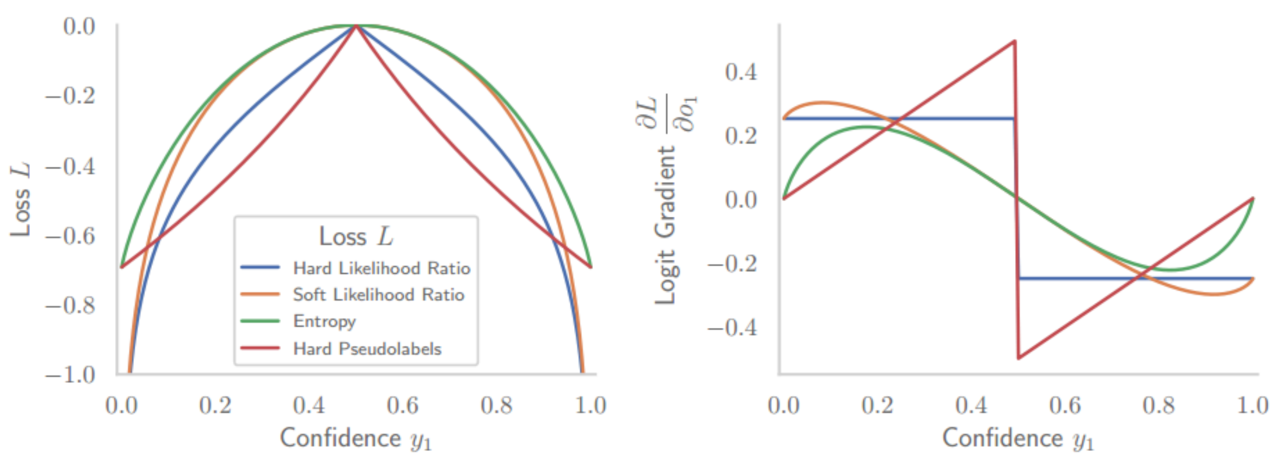
计算所有token的负对数似然作为熵的近似值：

$$H\left(\tilde{y}_j\right)=-\frac{1}{L_j} \sum_{k=1}^{L_j} \log P\left(r_j^k \mid t_j, r_j^{<k}\right),$$

使用标记数据的熵值θ百分位数（50%）作为阈值过滤低置信样本：

$$\mathcal{D}_{\text{selected}}=\left\{\left(t_j, \tilde{y}_j\right) \mid H\left(\tilde{y}_j\right) \leq \tau\right\}.$$

Universal Test-time Adaptation through Weight Ensembling, Diversity Weighting, and Prior Correction



熵最小化的梯度受低置信度预测的支配，使用软似然比损失（SLR）高置信度样本的梯度相对较大，加权过滤掉不多样不可靠的样本：

$$\mathcal{L}_{\mathrm{SLR}}(\hat{\boldsymbol{y}}_{ti}) = -\sum_{c} w_{ti}\, \hat{y}_{tic} \log\left(\frac{\hat{y}_{tic}}{\sum_{j \neq c} \hat{y}_{tij}}\right),$$

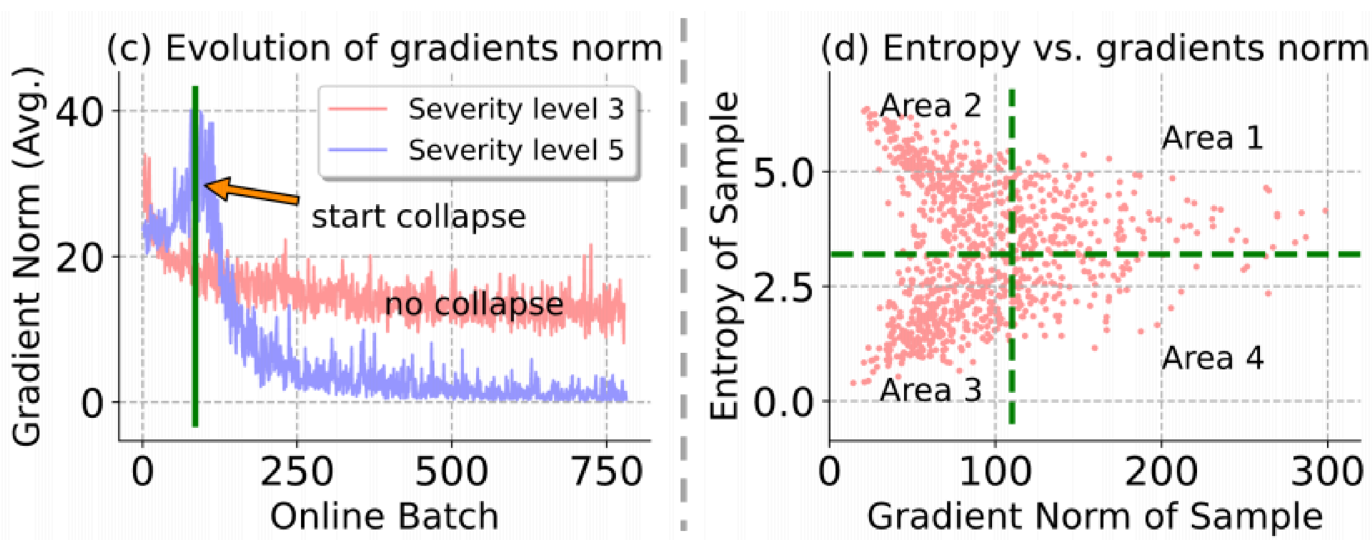
$$w_{\mathrm{div},ti} = 1 - \frac{\hat{\boldsymbol{y}}_{ti}^{\mathrm{T}}\, \bar{\boldsymbol{y}}_{t}}{\|\hat{\boldsymbol{y}}_{ti}\|\, \|\bar{\boldsymbol{y}}_{t}\|}.$$

$$w_{\mathrm{cert},ti} = -H(\hat{\boldsymbol{y}}_{ti}) = \sum_{c} \hat{y}_{tic} \log \hat{y}_{tic}.$$

$$\boldsymbol{w}_{t} = \exp\left(\frac{\boldsymbol{w}_{\mathrm{div},t}\, \boldsymbol{w}_{\mathrm{cert},t}}{\tau}\right).$$

Towards Stable Test-time Adaptation in Dynamic Wild World

除了高熵样本，某些样本产生的大梯度会使模型崩溃。除了过滤高熵样本，也要过滤大梯度



将模型更新到损失函数的平坦区域：

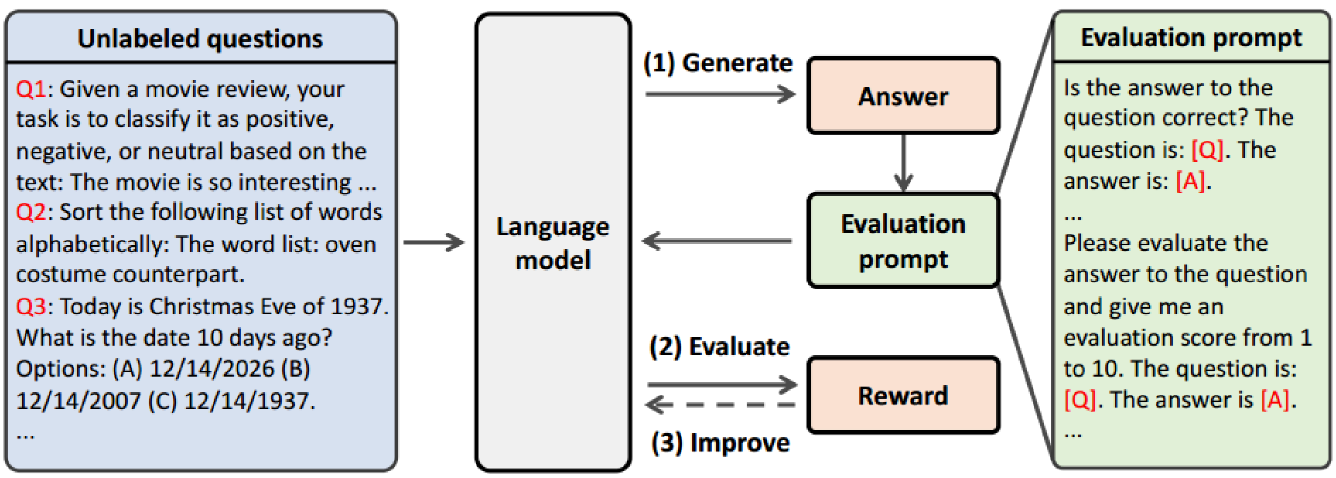
$$\min_{\Theta} E^{SA}(\mathbf{x};\Theta), \quad \text{where} \quad E^{SA}(\mathbf{x};\Theta) \triangleq \max_{\|\boldsymbol{\epsilon}\|_2 \leq \rho} E(\mathbf{x};\Theta + \boldsymbol{\epsilon}).$$

$$\hat{\epsilon}(\Theta) = \rho \,\mathrm{sign}\left(\nabla_\Theta E(\mathbf{x};\Theta)\right) \left|\nabla_\Theta E(\mathbf{x};\Theta)\right| / \left\|\nabla_\Theta E(\mathbf{x};\Theta)\right\|_2 .$$

$$\nabla_\Theta E^{SA}(\mathbf{x};\Theta) \approx \nabla_\Theta E(\mathbf{x};\Theta)\big|_{\Theta+\hat{\epsilon}(\Theta)} .$$

模型发生崩溃后熵很小，当低于阈值时重置模型参数为原始值。

LANGUAGE MODEL SELF-IMPROVEMENT BY REINFORCEMENT LEARNING CONTEMPLATION



unlabelled question输入LM，将得到的question和answer与Evaluation prompt一起输入LM得到对回答的评分，用评分作为奖励进行强化学习。