TS-LLM Enhances LLM Inference
Decoding by Tree-Search

Node Expansion

Sentence-level

1,1,3,8

8*3=24
(left: 1,1,24)

3+8=11
(left: 1,1,11)

1*24=24
(left: 1,24)

11+1=12
(left 1,12)

Token-level

1,1,3,8

8   3

*   +   −

3   1   3   1

Inference via AlphaZero-Like Tree-Search

Act: $a \sim \dfrac{N(s_t, a)^{1/\tau}}{\sum_b N(s_t, b)^{1/\tau}}$

$v_{s_{\text{leaf}}}$

→ Select
⇢ Backup
⬚ Expand
● Seleted node
● Eval leaf node
○ Visited node $\times N_{\text{simulation}}$

Policy improvement

Return

Training

Policy Distillation

Policy Learning Buffer

$L(\theta) = \mathrm{E}_{(\mathbf{x}^j, \mathbf{y}^j) \sim \mathcal{D}^+} \left[-\log \pi_\theta(\mathbf{y}^j | \mathbf{x}^j) \right]$

Policy Evaluation

Value Learning Buffer

Corrent Answers   Wrong Answers

Return
Estimation

$L(\phi) = \mathrm{E}_{\mathcal{D}} \left[\sum_{t=0}^{T^j-1} ||v_\phi(s_{0:t}^j | \mathbf{x}^j) - z_t^j||_2^2 \right]$

TS-LLM Provides Improved Paths
for LLM Training