Comprehensive Performance Analysis of Weight-of-Thought Reasoning
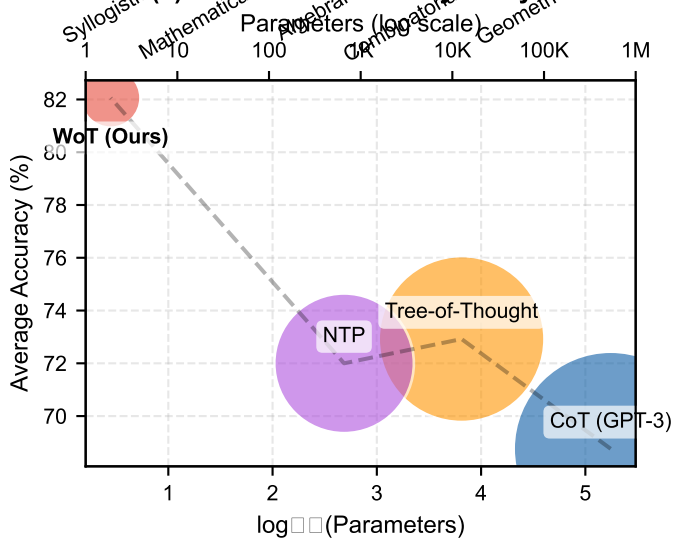
(a) Task-Specific Accuracy

Legend: CoT, Tree-of-Thought, NTP, WoT (Ours)

Syllogistic: 72.5, 76.8, 74.2, 84.3
Mathematical: 68.3, 72.5, 75.6, 86.5
Algebraic: 74.1, 78.3, 73.8, 82.7
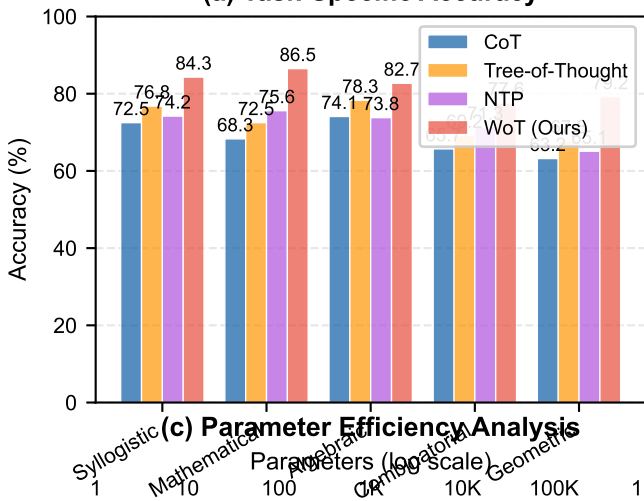
(b) Training Convergence Analysis

Legend: WoT (Ours), CoT, Tree-of-Thought
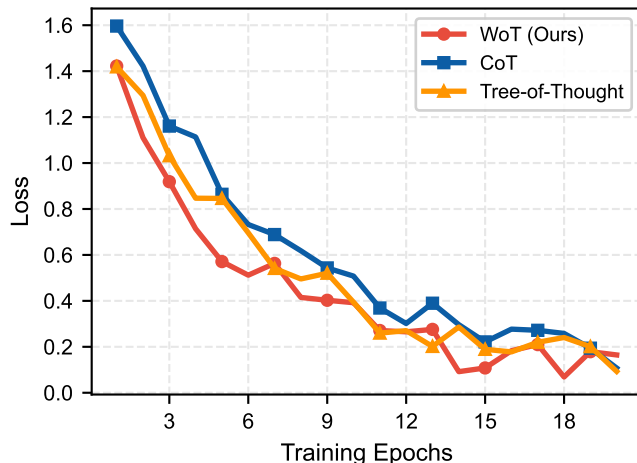
(c) Parameter Efficiency Analysis

WoT (Ours): 82
NTP
Tree-of-Thought
CoT (GPT-3)

x-axis: $\log_{10}$(Parameters)
y-axis: Average Accuracy (%)

(d) Reasoning Capability Radar

Legend: CoT, Tree-of-Thought, WoT (Ours)

Axes: Reasoning Depth, Branching Factor, Integration Score, Error Recovery