

GSM8Kにおけるエントロピー適応型 self-consistency サンプリングの分析

佐々木 駿^{1,a)} 川島 英之^{1,b)}

概要：本稿は、GSM8K に対して self-consistency を $n = 50$ まで拡張したときの利得とコストを精査し、一番手票率等の指標に基づく動的制御の必要性を示す。Meta-Llama-3-8B-Instruct を用いた推論では、厳密精度は 45.7% から 81.0% へ向上する一方、 $n = 20$ 以降の増分は小さく、レイテンシが線形に増加した。全 8,792 問に対する相関では、一番手票率が最も強く結び付き、二番手票率が負に相関した。Prefix エントロピー単体の識別力は限定的である。

1. 背景

Chain-of-Thought (CoT) プロンプティングは大規模言語モデルに逐次推論を促し、途中の思考過程を明示的に出力させることで論理的一貫性と正答率の向上を狙う手法である [1]。自明な例として「太郎はリンゴを 3 個持ち、花子はその 2 倍を持つ。合計はいくつか？」という問い合わせに対し、CoT なしでは“6”と誤答する一方、“step by step で考えて”と指示した CoT では「 $3 + (2 \times 3) = 9$ 」まで明示して正答に到達する。

self-consistency は同一問題に対して CoT を複数回独立生成し、得られた最終回答の多数決を採用する集約法である [2]。例えば“クッキーは 1 箱 12 個。3 箱買って 2 個食べた。残りは？”の問い合わせに対し、 $n = 3$ の self-consistency は [36, 36, 34] という回答列から 36 を選び、单一サンプル時の手順抜けを補正する。本研究では self-consistency を $n = 50$ まで拡張したときの利得とコストを精査する。

2. 実験設定

2.1 データセット

GSM8K[3] は小学生向け算数文章題 8,792 問からなるデータセットであり、各行が “question” と “answer” を含む JSON Lines 形式で提供される。例として “Sarah has 3 apples and buys 2 more. How many apples does she have now?” に対し正解 “5” が付与される。学習は行わず、推論結果の最終数値を抽出して strict accuracy で採点した。

¹ 慶應義塾大学
Keio University
a) shun.sasaki@keio.jp
b) river@sfc.keio.ac.jp

2.2 環境

Meta-Llama-3-8B-Instruct を Hugging Face Transformers で実行し、NVIDIA H100 (80GB) を 8 基用いた。推論パラメータは temperature=0.7, top-p=0.9, max_new_tokens=256, min_new_tokens=1 とし、プロンプト文 “You are a helpful math tutor. Solve the following problem step by step and provide only the final numeric answer.” を固定で使用した。

2.3 Self-consistency の設定

各問題について $n = 50$ サンプルを生成（シード固定）し、得られた 50 本の思考列と最終回答を保存した。先頭 $n \in \{1, 3, 5, 7, 10, 15, 20, 30, 40, 50\}$ に対する多数決をポストホックで再計算し、prefix entropy ($K = 32$)、vote margin、second vote ratio、サンプルごとの生成時間、各 n の strict / normalized accuracy を記録した。

3. 分析

3.1 精度とレイテンシ

self-consistency を $n = 50$ まで拡張すると GSM8K の厳密精度は 45.7% から 81.0% へ伸びるが、 $n = 20$ 以降の増分は小さく、レイテンシが線形に増加する。高コスト帯では一律にサンプルを増やすのではなく、指標に基づく動的制御が不可欠である。図 1 に精度推移を、表 1 に代表サンプル数の厳密精度と推定レイテンシを示す。 $n \leq 10$ で急激に精度が向上し、 $n \geq 20$ では増分が緩やかな漸近挙動を示す。多数決によるばらつき抑制は初期サンプルで大きく寄与し、高サンプル域ではコストに対して改善が頭打ちになる。

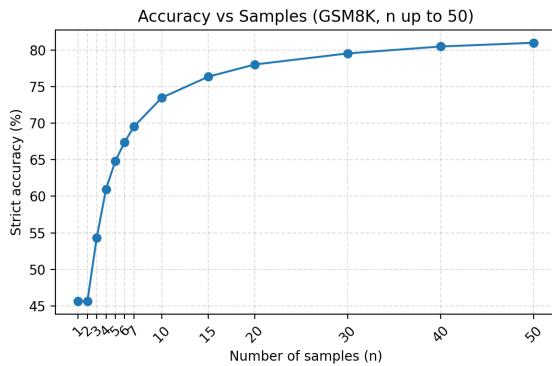


図 1 Self-consistency サンプル数に対する厳密精度の推移

表 1 代表サンプル数における厳密精度と推定レイテンシ

n	精度 (strict)	レイテンシ [s]
1	0.457	0.39
5	0.648	1.97
10	0.735	3.94
20	0.780	7.89
30	0.795	11.83
50	0.810	19.72

表 2 正答率と主要指標の相関 (全 8,792 問)

指標	相関係数
一番手票率	+0.37
二番手票率	-0.16
Prefix エントロピー	+0.08
生成時間	-0.10

3.2 信頼度指標と追加票の挙動

先に各指標の定義を整理する。

- **一番手票率:** 最多数回答の得票比率。0.5 で拮抗、1.0 で全会一致。
- **二番手票率:** 準多数回答の得票比率。高いほど票が割れている。
- **Prefix エントロピー:** 先頭 32 ワードの平均トーカンエントロピー。
- **生成時間:** 50 サンプル生成に要した合計時間。

表 2 に各指標と最終正答率の相関係数を示す。一番手票率が最も強く結び付き、二番手票率が負の相関を示す一方、prefix エントロピー単体の識別力は限定的である。図 2 ではサンプル数を増やした際の改善・劣化件数を可視化した。初期サンプルでは改善件数が劣化件数を明確に上回る一方、 $n \geq 20$ では両者が接近し僅差問題の劣化が残存することが読み取れる。中盤以降も改善が優勢だが、僅差領域では逆転劣化が生じ得るため票率を中心とした動的停止戦略が必要だと考えられる。

4. 結論

Self-consistency のサンプル数を最大 50 まで拡張すると GSM8K の厳密精度は 45.7% から 81.0% へ向上するが、

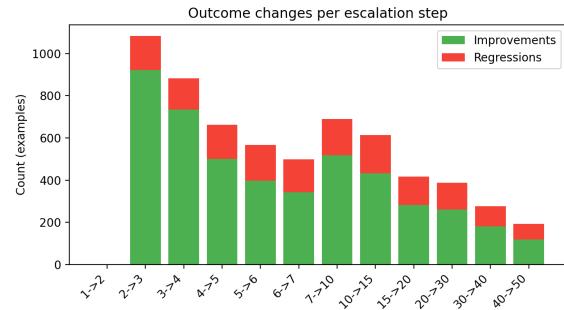


図 2 サンプル数増加ごとの改善・劣化件数の推移

$n = 20$ 以降は伸びが鈍化しレイテンシが急増する。一番手票率と二番手票率が最終正答率と最も強く結び付き、prefix エントロピー単体の識別力は限定的である。サンプル生成時間が長い問題ほど誤答が多く、コスト指標と信頼度指標を両立させた動的制御が必要である。

謝辞 本論文は、NEDO(国立研究開発法人新エネルギー・産業技術総合開発機構)の委託事業「ポスト 5G 情報通信システム基盤強化研究開発事業」(JPNP20017) および JPNNP16007、JSPS 科研費 25H00446、JST CREST JPMJCR24R4、セコム科学技術財団、JST COI-NEXT SQAI (JPMJPF2221)、JST ムーンショット型研究開発事業 JPMJMS2215 の支援を受けたものです。

参考文献

- [1] Wei, J., Wang, X., Schuurmans, D., Bosma, M., Chi, E., Xia, F., Le, Q., Zhou, D.: “Chain of Thought Prompting Elicits Reasoning in Large Language Models,” arXiv:2201.11903, 2022.
- [2] Wang, X., Wei, J., Schuurmans, D., Le, Q., Chi, E., Zhou, D.: “Self-Consistency Improves Chain of Thought Reasoning in Language Models,” arXiv:2203.11171, 2022.
- [3] Cobbe, K., Kosaraju, V., Bavarian, M., Chen, M., Jun, H., Kaiser, L., Plappert, M., Tworek, J., Hilton, J., Nakano, R., Hesse, C., Schulman, J.: “Training Verifiers to Solve Math Word Problems,” arXiv:2110.14168, 2021.