Novelty Assessment Report

Paper: Taming Imperfect Process Verifiers: A Sampling Perspective on Backtracking
PDF URL: https://openreview.net/pdf?id=MFDkLbcydi
Venue: ICLR 2026 Conference Submission
Year: 2026
Report Generated: 2025-12-30

Abstract

Test-time algorithms that combine the generative power of language models with process verifiers that assess the quality of partial generations offer a promising lever for eliciting new reasoning capabilities, but the algorithmic design space and computational scaling properties of such approaches are still opaque, and their benefits are far from apparent when one accounts for the cost of learning a high-quality verifier. Our starting point is the observation that seemingly benign errors in a learned verifier can lead to catastrophic failures for standard decoding techniques due to error amplification during the course of generation. We then ask: can this be improved with more sophisticated decoding strategies?

We introduce a new process-guided test-time sampling algorithm, VGB, which uses theoretically grounded backtracking to achieve provably better robustness to verifier errors. VGB interprets autoregressive generation as a random walk on a tree of partial completions, with transition probabilities guided by the process verifier and base model; crucially, backtracking occurs probabilistically. This process generalizes the seminal Sinclair-Jerrum random walk (Sinclair & Jerrum, 1989) from the literature on approximate counting and sampling in theoretical computer science, and a conceptual contribution of our work is to highlight parallels with this literature. Empirically, we demonstrate on both synthetic and real language modeling tasks that VGB outperforms baselines on a variety of metrics.

Core Task Landscape

This paper addresses: **Process-Guided Language Model Generation with Imperfect Verifiers**

A total of **5 papers** were analyzed and organized into a taxonomy with **6 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Process Verification and Value Modeling**
- **Test-Time Decoding with Process Guidance**
- **Post-Training Optimization with Imperfect Feedback**
- **Domain-Specific Applications of Imperfect Verification**

Complete Taxonomy Tree

- Process-Guided Language Model Generation with Imperfect Verifiers Survey Taxonomy
- Process Verification and Value Modeling
  - Process Reward Model Training and Scaling (1 papers)
  - [5] Rewarding Progress: Scaling Automated Process Verifiers for LLM Reasoning (Setlur, 2024) View paper
  - Uncertainty-Aware Value Modeling (1 papers)
  - [4] Robust Search with Uncertainty-Aware Value Models for Language Model Reasoning (Yu Fei, 2025) View paper
- Test-Time Decoding with Process Guidance
  - Backtracking-Based Sampling for Verifier Robustness ★ (1 papers)
  - [0] Taming Imperfect Process Verifiers: A Sampling Perspective on Backtracking (Anon et al., 2026) View paper
- Post-Training Optimization with Imperfect Feedback
  - Adversarial Critic-Based Reinforcement Learning (1 papers)
  - [3] RLAC: Reinforcement Learning with Adversarial Critic for Free-Form Generation Tasks (WU Mian, 2025) View paper
- Domain-Specific Applications of Imperfect Verification
  - Video Question Answering with Imperfect Reasoning (1 papers)
  - [1] ReasVQA: Advancing VideoQA with Imperfect Reasoning Process (Liang Jian-xin, 2025) View paper
  - Constraint-Guided Code Generation with Semantic Verification (1 papers)
  - [2] InstructFlow: Adaptive Symbolic Constraint-Guided Code Generation for Long-Horizon Planning (H Chi, 2025) View paper

Narrative

Core task: process-guided language model generation with imperfect verifiers. The field addresses how to leverage intermediate verification signals—often noisy or unreliable—to steer language model outputs toward correct solutions. The taxonomy organizes work into several main branches: Process Verification and Value Modeling focuses on learning step-level reward or value functions that assess partial reasoning traces; Test-Time Decoding with Process Guidance explores inference-time strategies such as search and sampling that use these verifiers to select or refine outputs; Post-Training Optimization with Imperfect Feedback examines how to train models using noisy process signals; and Domain-Specific Applications of Imperfect Verification applies these techniques to specialized settings like mathematical reasoning or visual question answering. Representative works such as ReasVQA[1] and InstructFlow[2] illustrate how process supervision can be integrated into both training pipelines and domain-specific architectures, while methods like RLAC[3] and Rewarding Progress[5] highlight different strategies for handling imperfect feedback during optimization.

A central theme across these branches is the tension between exploiting verifier guidance and mitigating verifier errors. Many studies explore how to make decoding robust when step-level scores are unreliable, for instance by incorporating uncertainty estimates as in

Robust Search Uncertainty[4] or by designing search procedures that can recover from incorrect intermediate assessments. Taming Imperfect Verifiers[0] sits within the Test-Time Decoding branch, specifically addressing backtracking-based sampling for verifier robustness. Its emphasis on allowing the generation process to revisit and correct earlier steps aligns closely with the broader goal of making search resilient to noisy signals, a concern shared by works like Robust Search Uncertainty[4]. Compared to approaches that primarily refine verifier training or aggregate multiple signals, Taming Imperfect Verifiers[0] focuses on the inference-time mechanism itself, offering a complementary perspective on how to navigate imperfect guidance without requiring perfect step-level supervision.

Related Works in Same Category

No sibling papers and no sibling subtopics were found under the same parent taxonomy node; the paper appears structurally isolated in the taxonomy.

Contributions Analysis

Overall novelty summary. The paper introduces VGB, a value-guided sampling algorithm with stochastic backtracking designed to mitigate error amplification from imperfect process verifiers during autoregressive generation. Within the taxonomy, it occupies the 'Backtracking-Based Sampling for Verifier Robustness' leaf under 'Test-Time Decoding with Process Guidance'. Notably, this leaf contains only the original paper itself, with no sibling papers identified in the taxonomy. This suggests the specific focus on probabilistic backtracking for verifier robustness represents a relatively sparse research direction within the broader field of process-guided generation.

The taxonomy reveals that neighboring work primarily addresses verifier training (Process Reward Model Training, Uncertainty-Aware Value Modeling) and post-training optimization (Adversarial Critic-Based RL), rather than test-time decoding strategies. The closest related direction is 'Uncertainty-Aware Value Modeling', which tackles verifier imperfection through uncertainty quantification in value functions, whereas VGB addresses the same challenge through inference-time backtracking. Domain-specific applications (video QA, code generation) apply process guidance to specialized tasks but do not focus on the core algorithmic robustness question that VGB targets.

Among the three contributions analyzed, none were clearly refuted by the fourteen candidates examined. The VGB algorithm itself was compared against four candidates with no refutations found. The theoretical analysis of error amplification examined one candidate, and the connection to approximate sampling theory examined nine candidates, both without finding overlapping prior work. These statistics suggest that within the limited search scope—top-K semantic matches plus citation expansion—the paper's contributions appear distinct from existing literature, though the small candidate pool (fourteen total) means this assessment is necessarily preliminary.

Based on the limited literature search, the work appears to occupy a novel position at the intersection of test-time decoding and verifier robustness. The absence of sibling papers in its taxonomy leaf and the lack of refutations across fourteen candidates suggest originality, though a more exhaustive search covering broader decoding strategies and approximate sampling methods would strengthen this assessment. The taxonomy structure indicates the paper addresses a gap between verifier training methods and their deployment at inference time.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: VGB: Value-Guided Sampling with Stochastic Backtracking algorithm

Description: The authors propose VGB, a novel algorithm that interprets autoregressive generation as a random walk on a tree of partial generations with probabilistic backtracking. The algorithm generalizes the Sinclair-Jerrum random walk and provides theoretical guarantees for robustness to value function errors during guided generation.

This contribution was assessed against **4 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. SequenceMatch: Imitation Learning for Autoregressive Sequence Modelling with Backtracking

URL: View paper

Brief Assessment

SequenceMatch[8] focuses on imitation learning for autoregressive sequence modeling with backtracking in text generation contexts, not on value-guided sampling with theoretical guarantees for robustness to value function errors in general RL frameworks as proposed in the original paper.

---

2. Large Language Model-Driven Multi-agent Collaborative Framework for Chinese Grammatical Error Correction

URL: View paper

Brief Assessment

Multi-agent Grammar Correction[7] focuses on Chinese grammatical error correction using multi-agent collaboration, not value-guided sampling with backtracking for autoregressive generation robustness in language models.

---

3. ROCODE: Integrating Backtracking Mechanism and Program Analysis in Large Language Models for Code Generation

URL: View paper

Brief Assessment

ROCODE[9] focuses on code generation with backtracking for error correction during generation, not on value-guided sampling for autoregressive generation with process verifiers. The backtracking mechanisms serve fundamentally different purposes in different problem domains.

---

4. Language model uncertainty quantification with attention chain

URL: View paper

Brief Assessment

Attention Chain Uncertainty[6] focuses on uncertainty quantification for language model outputs using attention-based token selection, not on value-guided sampling with backtracking for autoregressive generation robustness as in VGB.

Contribution 2: Theoretical analysis of error amplification in action-level sampling

Description: The authors provide theoretical examples showing that standard action-level rejection sampling catastrophically amplifies seemingly benign errors in learned value functions across long generation horizons, motivating the need for more sophisticated decoding strategies.

This contribution was assessed against **1 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Scalable Offline Model-Based RL with Action Chunks

URL: View paper

Brief Assessment

Action Chunks[20] focuses on action-chunk models for offline model-based RL to reduce compounding errors in dynamics models, not on theoretical analysis of error amplification in action-level rejection sampling with approximate value functions as in the original paper.

Contribution 3: Connection between value-guided inference and approximate sampling theory

Description: The authors establish conceptual connections between value-guided language model inference and classical approximate counting and sampling techniques from theoretical computer science, particularly the Sinclair-Jerrum random walk, opening avenues for transferring ideas between these areas.

This contribution was assessed against **9 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

---

1. Hypothesis-driven theory-of-mind reasoning for large language models

URL: View paper

Brief Assessment

Hypothesis-driven ToM[10] focuses on theory-of-mind reasoning using sequential Monte Carlo for mental state inference in social contexts, not on value-guided language model inference or approximate sampling for test-time alignment as in the original paper.

---

2. Quasi-random Multi-Sample Inference for Large Language Models

URL: View paper

Brief Assessment

Quasi-random Multi-Sample[14] focuses on arithmetic sampling for diverse multi-sample generation in LLM decoding tasks (reasoning, translation), not on establishing theoretical connections between value-guided inference and approximate counting/sampling techniques from theoretical computer science.

---

3. Inference-aware fine-tuning for best-of-n sampling in large language models

URL: View paper

Brief Assessment

Inference-aware Fine-tuning[12] focuses on best-of-n sampling strategies for LLMs using verifiers, not on establishing connections to classical approximate counting/sampling theory like the Sinclair-Jerrum random walk. The candidate addresses a different problem domain (inference-time optimization for LLMs) without engaging with the theoretical computer science foundations that the original contribution claims.

---

4. PILAF: Optimal Human Preference Sampling for Reward Modeling

URL: View paper

Brief Assessment

PILAF[18] focuses on preference sampling strategies for reward modeling in RLHF, not on value-guided language model inference or approximate sampling theory from theoretical computer science.

---

5. Coarticulatory inference propagation in probabilistic attention meshes for large language model sampling flux stabilization

URL: View paper

Brief Assessment

Coarticulatory Inference[15] focuses on probabilistic attention mechanisms for sampling flux stabilization in LLMs, not on establishing connections between value-guided inference and classical approximate counting/sampling techniques like the Sinclair-Jerrum random walk.

---

6. Accelerating large language model decoding with speculative sampling

URL: View paper

Brief Assessment

Speculative Sampling[16] focuses on accelerating transformer decoding through parallel scoring of draft continuations, not on establishing theoretical connections between value-guided inference and approximate sampling/counting techniques from theoretical computer science.

---

7. Out-of-Vocabulary Sampling Boosts Speculative Decoding

URL: View paper

Brief Assessment

OOV Speculative Decoding[13] focuses on vocabulary pruning and token sampling for speculative decoding in language models, not on value-guided inference or approximate sampling theory from theoretical computer science. The papers address fundamentally different problems.

---

8. Large language monkeys: Scaling inference compute with repeated sampling

URL: View paper

Brief Assessment

Language Monkeys[11] focuses on repeated sampling from language models to improve coverage and does not discuss approximate sampling theory from theoretical computer science or connections to the Sinclair-Jerrum random walk. The paper's approach is empirically-driven rather than theoretically grounded in classical approximate counting techniques.

---

9. Enabling Approximate Joint Sampling in Diffusion LMs

URL: View paper

Brief Assessment

Approximate Joint Sampling[17] focuses on diffusion language models and joint token sampling, not on value-guided inference for language models or connections to Sinclair-Jerrum random walks for approximate counting/sampling.

Appendix: Text Similarity Detection

No high-similarity text segments were detected across any compared papers.

References

- [0] Taming Imperfect Process Verifiers: A Sampling Perspective on Backtracking View paper
- [1] ReasVQA: Advancing VideoQA with Imperfect Reasoning Process View paper
- [2] InstructFlow: Adaptive Symbolic Constraint-Guided Code Generation for Long-Horizon Planning View paper
- [3] RLAC: Reinforcement Learning with Adversarial Critic for Free-Form Generation Tasks View paper
- [4] Robust Search with Uncertainty-Aware Value Models for Language Model Reasoning View paper
- [5] Rewarding Progress: Scaling Automated Process Verifiers for LLM Reasoning View paper
- [6] Language model uncertainty quantification with attention chain View paper
- [7] Large Language Model-Driven Multi-agent Collaborative Framework for Chinese Grammatical Error Correction View paper
- [8] SequenceMatch: Imitation Learning for Autoregressive Sequence Modelling with Backtracking View paper
- [9] ROCODE: Integrating Backtracking Mechanism and Program Analysis in Large Language Models for Code Generation View paper
- [10] Hypothesis-driven theory-of-mind reasoning for large language models View paper
- [11] Large language monkeys: Scaling inference compute with repeated sampling View paper
- [12] Inference-aware fine-tuning for best-of-n sampling in large language models View paper
- [13] Out-of-Vocabulary Sampling Boosts Speculative Decoding View paper
- [14] Quasi-random Multi-Sample Inference for Large Language Models View paper
- [15] Coarticulatory inference propagation in probabilistic attention meshes for large language model sampling flux stabilization View paper
- [16] Accelerating large language model decoding with speculative sampling View paper
- [17] Enabling Approximate Joint Sampling in Diffusion LMs View paper
- [18] PILAF: Optimal Human Preference Sampling for Reward Modeling View paper
- [19] Approximately Aligned Decoding View paper
- [20] Scalable Offline Model-Based RL with Action Chunks View paper