Self-Perturbation Learning for Mathematical Reasoning Verification

Kreasof AI
Preprint · Under Review

Abstract

We propose **Self-Perturbation Learning (SPL)**, a self-supervised framework for training verifier models to assess the correctness and quality of mathematical reasoning. SPL introduces context-aware "impostor" elements into training data by substituting words with semantically similar alternatives, guided by cosine distances in embedding space. Using a modified ModernBERT architecture and the AutoMathText dataset (30K samples), we demonstrate that SPL-trained models achieve robust performance in distinguishing high-quality mathematical content from irrelevant text, logical fallacies, and subtle errors. Our revised sequence quality scoring method produces interpretable outputs (range: -1 to 1), with high scores for valid reasoning (e.g., 0.85) and negative scores for irrelevant sequences (e.g., -153). This work lays the foundation for scalable, domain-specific verification systems without reliance on manual labelling.

Keywords: Self-supervised learning, mathematical reasoning, verifier models, error detection, ModernBERT

1. Introduction

Large language models (LLMs) struggle with mathematical consistency, often producing plausible but incorrect reasoning steps ("hallucinations"). Existing verification methods rely on rule-based systems or costly human annotation, limiting scalability. We address this gap with **SPL**, a self-supervised approach that trains models to detect synthetic errors ("impostors") generated through embedding-guided perturbations. By fine-tuning ModernBERT on the AutoMathText corpus, we create a verifier that quantifies reasoning quality while requiring no manual labels.

2. Methodology

2.1 Self-Perturbation Learning (SPL)

- **Impostor Generation**: For each token, substitute it with a semantically similar "impostor" word sampled from the top 150 frequent AutoMathText terms. Substitutions are ranked by cosine similarity in ModernBERT's embedding space, with three difficulty levels:
  - **Hard**: Top 10 nearest neighbours.
  - **Medium**: Neighbors ranked 10–50.
  - **Easy**: Neighbors ranked 50–100.
- **Training Objective**: A binary classification task (original vs. impostor tokens) with BCE loss.

2.2 Sequence Quality Scoring

The verifier computes a **negative log-likelihood** of token-level impostor probabilities. We apply min-max scaling to map scores into an interpretable range:

$$[\text{Revised Score} = \frac{-\text{NLL} - \min}{\max - \min}]$$

where $\min = -1, \max = 0,$ producing scores from -1 (irrelevant) to 1 (high-quality).

2.3 Experimental Setup

- **Model**: ModernBERT-base (8K context) fine-tuned on 30K samples from AutoMathText's web-0.80-to-1.00 subset.
- **Training**: 1 epoch, batch size 16, AdamW (lr = 1e-5), NVIDIA GPU.

3. Results (30K Sample)

Text Type	Example	Revised Score
Irrelevant Sequence	“Cat sat on the mat”	-153.23
High-Quality Math Explanation	Algebraic simplification steps	0.85
Logical Fallacy	Invalid “Proof 1 = 2”	0.36–0.45
Correct Word Problem Answer	Natalia’s clip sales (72 total)	0.80
Incorrect Word Problem Answer	Betty’s irrelevant calculation	0.55

Key Findings:
1. The model reliably flags irrelevant text (scores < 0).
2. Logical fallacies receive intermediate scores, indicating sensitivity to flawed reasoning.
3. Correct answers score significantly higher than incorrect ones ($\Delta = 0.25$).

4. Discussion

Advantages of SPL

- **Self-Supervision**: Eliminates dependency on labelled data.
- **Interpretability**: Scores align with human intuition (e.g., -153 vs. 0.85).
- **Generality**: Framework applicable to code, scientific text, and multimodal data.

Limitations and Future Work

1. **Quantitative Benchmarks**: Pending AUC-ROC/precision-recall metrics for error detection.
2. **Structural Perturbations**: Current impostors focus on word substitutions; future work will perturb mathematical operators or step ordering.
3. **Scale**: Training on 500K+ samples (in progress) to improve robustness.

5. Conclusion

SPL demonstrates promising results in mathematical reasoning verification, achieving self-supervised, interpretable quality assessment. The 30K-sample model successfully differentiates valid reasoning from irrelevant or flawed content, providing a foundation for trustworthy AI systems in education and research. Updates with larger-scale training (500K samples) and structural perturbations will follow.

References

1. Zhang et al. (2024). *AutoMathText: Autonomous Data Selection with Language Models for Mathematical Texts*. arXiv:2402.07625.
2. Warner et al. (2024). *ModernBERT: Efficient Long-Context Fine-Tuning*. arXiv:2412.13663.

Code and Data Availability

- Code: https://github.com/kreasof-ai/self-perturbation-learning
- Data: Subset of math-ai/AutoMathText (Hugging Face Datasets)