BlueMO: A Comprehensive Collection of Challenging Mathematical Olympiad Problems from the Little Blue Book Series

Yizhou Chen   Yifan Luo   Yifan Zhang   Yang Yuan

Abstract

We introduce BlueMO, a comprehensive and challenging dataset of over 3,000 mathematical olympiad problems, meticulously curated from the esteemed "Little Blue Book" series (Second Edition). This series is a vital resource for top students in China training for national and international math competitions. BlueMO provides problems and detailed, human-expert solutions in a machine-readable format, covering a wide range of advanced topics including number theory, combinatorics, and geometry. Designed to advance and assess sophisticated, multi-step reasoning in Large Language Models (LLMs), this dataset serves as a rigorous new benchmark and a valuable training resource for high-level mathematical problem-solving in AI. Its origin from a non-English source also introduces novel problem structures and reasoning styles, enriching the landscape of available resources for the global research community.

1   Introduction

Advanced mathematical reasoning remains one of the most challenging frontiers for Large Language Models (LLMs). While models have shown remarkable progress on elementary and high-school-level math problems, their ability to perform complex, multi-step logical inference—a hallmark of human intelligence—is still limited. To spur progress, the AI research community requires high-quality, complex benchmark datasets that push the boundaries of current models for both training and evaluation.

Existing datasets like GSM8K (Cobbe et al., 2021) and MATH (Hendrycks et al., 2021) have been instrumental, but they primarily cover standard curriculum mathematics. To truly test and develop sophisticated reasoning, we need problems that demand creativity, abstract thinking, and the synthesis of multiple mathematical concepts, characteristic of mathematics olympiads.

The "Little Blue Book" series, published by East China Normal University Press, is a cornerstone resource for students in China striving to excel in mathematical olympiads. Renowned for its depth, challenging problems, and elegant, detailed solutions, the series spans critical domains, including Sets, Functions, Geometry, Number Theory, Graph Theory, and Combinatorics, providing rigorous

training for the highest levels of competition.

We introduce **BlueMO**, a new dataset meticulously curated from 14 volumes of this esteemed series (Second Edition) (Chen et al., 2024). By structuring over 2,100 complex problems and their human-written solutions into a machine-readable format, we provide a new, challenging resource aimed at pushing the boundaries of automated reasoning. The dataset's novelty lies in its difficulty, the high quality of its solutions, and its origin, offering problems with unique logical structures not commonly found in existing English-centric datasets.

2 Comparison with Existing Datasets

To position BlueMO in the landscape of mathematical reasoning datasets, we compare it with several prominent benchmarks in Table 1. BlueMO distinguishes itself through its focus on Olympiad-level difficulty, the inclusion of expert-written, step-by-step solutions, and its specialized subject matter. Unlike datasets focused on final-answer correctness, BlueMO's detailed solutions enable fine-grained evaluation of reasoning paths.

Table 1 Comparison of BlueMO with other mathematical reasoning datasets.

Dataset	Difficulty Level	Solution Style	Content Domain
BlueMO (Ours)	**National Olympiad**	**Full step-by-step human solutions**	**Advanced topics (Number Theory, Combinatorics, etc.)**
GSM8K	Grade School	Final answer with simple steps	Arithmetic, elementary algebra
MATH	High School / AIME	Final answer, some with steps	Standard high school curriculum
AIME	Olympiad (AIME)	Final answer only	Olympiad-style algebra, geometry, number theory

3 Dataset Composition and Curation

BlueMO encompasses a total of 14 volumes from the third edition of the "Little Blue Book" series, covering a wide range of mathematical topics primarily for the high school olympiad level. The dataset composition is detailed in Table 2.

3.1 Data Collection and Curation Process

The creation of BlueMO involved a multi-stage process to ensure the highest fidelity to the original source material.

1. **Manual Transcription:** All problems, solutions, and remarks were manually transcribed from the physical books by individuals with a strong background in mathematics. This approach was chosen over OCR to avoid errors in complex LaTeX formulas and symbolic notation.

2. **LaTeX Conversion:** The transcribed content was converted into clean, compilable LaTeX format. Special attention was paid to preserving the original structure of equations, proofs, and logical steps.

3. **Verification and Proofreading:** Each entry underwent a two-step verification process. First, an automated script checked for common LaTeX compilation errors. Second, each problem and solution was manually proofread by a separate team member to verify mathematical correctness and consistency with the source.

4. **Structuring and JSON Export:** The final, verified LaTeX content was parsed and structured into the JSON format described in Section 5, making it readily accessible for machine learning applications.

Table 2 BlueMO Dataset Volumes (High School Collection)

Original Volume Title (Chinese)	English Translation / Content Domain
小蓝书高中卷1 集合	Little Blue Book High School Vol.1: Sets
小蓝书高中卷2 函数与函数方程	Little Blue Book High School Vol.2: Functions & Functional Equations
小蓝书高中卷3 三角函数	Little Blue Book High School Vol.3: Trigonometric Functions
小蓝书高中卷4 平均值不等式与柯西不等式	Little Blue Book High School Vol.4: Mean Value & Cauchy Inequalities
小蓝书高中卷5 不等式的解题方法与技巧	Little Blue Book High School Vol.5: Methods for Solving Inequalities
小蓝书高中卷6 数列与数学归纳法	Little Blue Book High School Vol.6: Sequences & Mathematical Induction
小蓝书高中卷7 平面几何	Little Blue Book High School Vol.7: Plane Geometry
小蓝书高中卷8 复数与向量	Little Blue Book High School Vol.8: Complex Numbers & Vectors
小蓝书高中卷9 几何不等式	Little Blue Book High School Vol.9: Geometric Inequalities
小蓝书高中卷10 数论	Little Blue Book High School Vol.10: Number Theory
小蓝书高中卷11 组合数学	Little Blue Book High School Vol.11: Combinatorics
小蓝书高中卷12 图论	Little Blue Book High School Vol.12: Graph Theory
小蓝书高中卷13 组合极值	Little Blue Book High School Vol.13: Extremal Combinatorics
小蓝书高中卷14 高中数学竞赛中的解题方法与策略	Little Blue Book High School Vol.14: Problem-Solving Methods

4 Dataset Analysis

The dataset contains 2,110 problems distributed across the 14 volumes. The problems are inherently challenging, designed to test deep conceptual understanding rather than rote calculation. The solutions are particularly valuable, with an average length of over 150 words and often involving

multiple logical steps, case analyses, and creative constructions. This richness provides a strong signal for training models to produce detailed, human-like reasoning chains.

5 Data Structure and Usage

We provide both the raw data (`*.tex`) and the processed dataset in JSON format. The dataset is categorized by problem type, such as calculation, proof, etc., and includes references to any figures the problems use.

The JSON fields are defined as follows:

source_file: Path to the original `.tex` source file containing this problem.

problem_type: Problem category (e.g., "calculation", "proof", "find the value").

problem: Complete problem statement in LaTeX format, including any sub-questions.

solution: The detailed, step-by-step solution with mathematical derivations in LaTeX.

remark: Additional notes or comments related to the problem or its solution (empty if none).

figures: An array listing any associated diagram or image files (empty if none).

An example entry from the dataset is shown in Listing 1.

```json
{
  "source_file": "vol1_sets/chapter1/problem_1.tex",
  "problem_type": "calculation",
  "problem": "Example 1. Let the set be $M=\\left\\{x \\mid \\frac{ax-5}{x^2-a} <
    0, x \\in \\mathbb{R}\\right\\}$.\\n(1) When $a=4$, simplify the set $M$.\\n(2)
     If $3 \\in M$ and $5 \\notin M$, find the range of the real number $a$.",
  "solution": "Analysis: To simplify set $M$, we must solve the inequality $\\frac
    {ax-5}{x^2-a} < 0$.\\nSolution: (1) When $a=4$, we have\\n$$\\frac{4x-5}{x^2-4}
     < 0$$\\nwhich is equivalent to\\n$$\\left(x-\\frac{5}{4}\\right)(x+2)(x-2) <
    0.$$\\nThe solution is $x < -2$ or $\\frac{5}{4} < x < 2$.\\nTherefore, $M = \\
    left\\{x \\mid x < -2 \\text{ or } \\frac{5}{4} < x < 2\\right\\}$.\\n(2) Since
     $3 \\in M$, we have $\\frac{3a-5}{3^2-a} < 0$, which is $\\frac{3a-5}{9-a} < 0
    $.\\nThis simplifies to $(3a-5)(a-9) > 0$, so $a < \\frac{5}{3}$ or $a > 9$.\\
    nSince $5 \\notin M$, we have either $\\frac{5a-5}{5^2-a} \\ge 0$ or the
    denominator is zero ($5^2-a=0$).\\nThe inequality $\\frac{5(a-1)}{25-a} \\ge 0$
     combined with $a=25$ gives $1 \\le a \\le 25$.\\nCombining both conditions for
     $a$, we get the final range: $a \\in \\left[1, \\frac{5}{3}\\right) \\cup (9,
    25]$.",
  "remark": "Note: The condition that $5 \\notin M$ includes the case where the
    denominator $x^2-a$ becomes zero when $x=5$. This possibility is easy to
    overlook.",
  "figures": []
}
```

Listing 1 Example of a JSON data entry in BlueMO.

6 Potential Usages

This dataset is a valuable resource for AI researchers and developers, with several key applications:

- **Training & Fine-Tuning:** Enhancing LLMs' capabilities in advanced mathematical reasoning. The detailed, high-quality solutions provide an excellent basis for supervised fine-tuning or reinforcement learning from human feedback (RLHF) to teach models complex, multi-step derivation.

- **AI Evaluation:** Serving as a challenging benchmark to test the problem-solving proficiency and logical rigor of state-of-the-art AI systems. Evaluation can be performed not only on the final answer but also on the correctness of the reasoning path.

- **Formal Verification:** Providing a rich source of problems for formalization into theorem provers (e.g., Lean, Isabelle, Coq). This can be used to evaluate an AI's ability to bridge the gap between informal and formal mathematical reasoning.

- **Comparative Analysis:** Enabling systematic assessment of reasoning skills across different models, prompting strategies, and methodologies. The dataset's difficulty makes it ideal for differentiating the capabilities of top-performing models.

7 License

The BlueMO dataset is licensed under a **Creative Commons Attribution-NoDerivatives 4.0 International License (CC BY-ND 4.0)**. This means you are free to share and use the dataset for any purpose, including commercially, as long as you give appropriate credit and do not distribute modified versions of it. The full license details are available at https://creativecommons.org/licenses/by-nd/4.0/.

8 Citation

If you use the BlueMO dataset in your research, please consider citing it as follows:

```
@misc{bluemo2024,
    title={BlueMO: A comprehensive collection of challenging mathematical
  olympiad problems from the little blue book series},
    author={Chen, Yizhou and Luo, Yifan and Zhang, Yifan and Yuan, Yang},
    year={2024}
}
```

Listing 2 BibTeX entry for the BlueMO dataset.

References

Yizhou Chen, Yifan Luo, Yifan Zhang, and Yang Yuan. Bluemo: A comprehensive collection of challenging mathematical olympiad problems from the little blue book series, 2024.

Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, et al. Training verifiers to solve math word problems. *arXiv preprint arXiv:2110.14168*, 2021.

Dan Hendrycks, Collin Burns, Saurav Kadavath, Akul Arora, Steven Basart, Eric Tang, Dawn Song, and Jacob Steinhardt. Measuring mathematical problem solving with the math dataset. *arXiv preprint arXiv:2103.03874*, 2021.