上海交通大学
SHANGHAI JIAO TONG UNIVERSITY

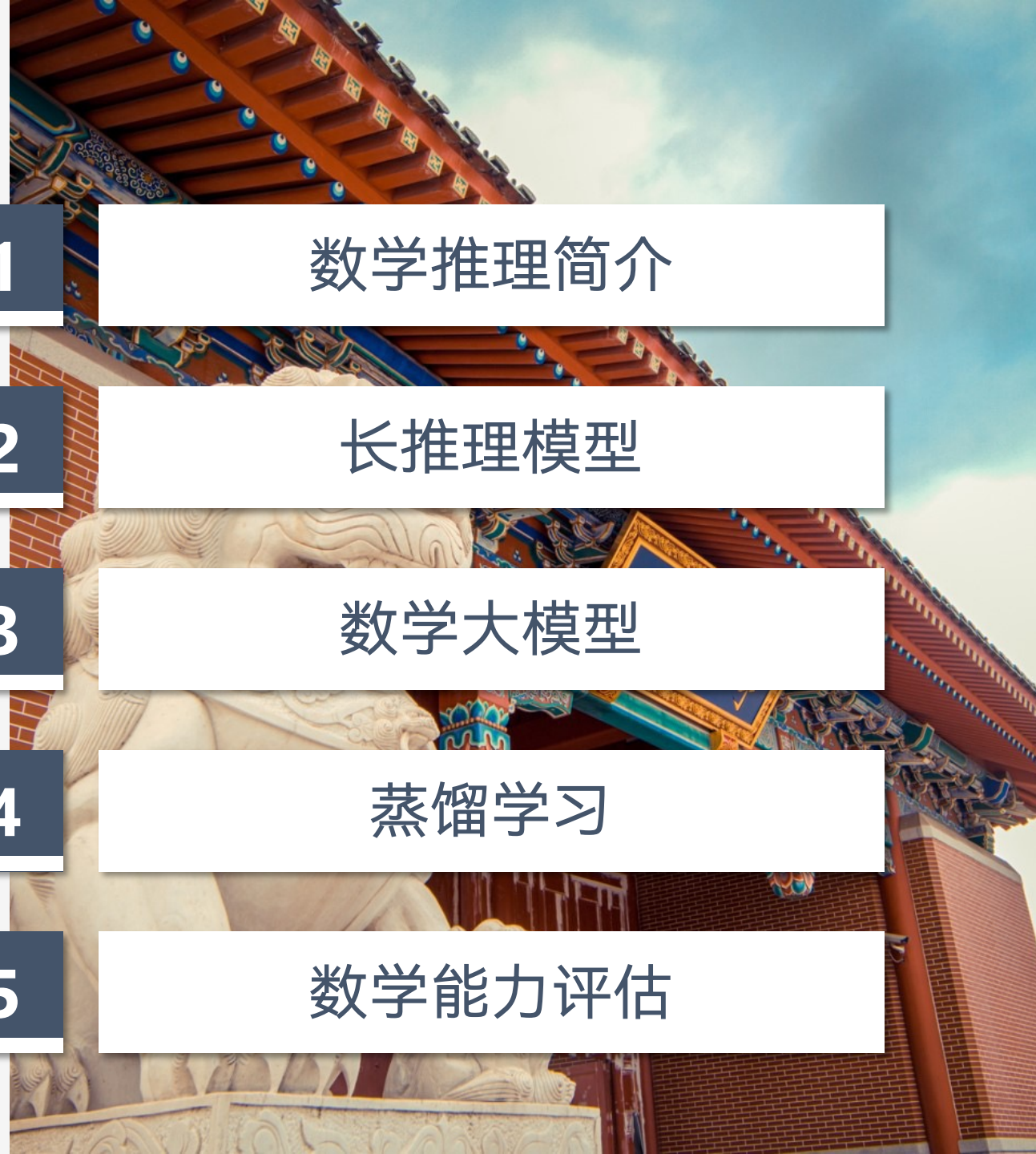
SJTU

LLM数学推理

陈星宇

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

目 录



数学推理简介

- **数学推理**：LLM通过逻辑分析、符号运算与多步推导解决数学问题的能力，需理解数学概念（如代数、几何）并生成正确答案
- **关键能力**：
  - 多步推理：分解复杂问题为子步骤
  - 符号理解：处理抽象符号与公式
  - 数学运算：进行精准的计算
- **应用场景**：数学竞赛解题、教育辅导、科学研究中的公式推导

如何在LLM中实现数学推理？

- **思维链[1]（CoT）**：显式生成推理步骤
- **树搜索**：蒙特卡洛树搜索[2]：结合过程奖励模型（PRM）进行不同思路的探索
- **自洽性[3]（Self-Consistency）**：多次采样不同推理路径，投票选择高频答案以降低错误
- **工具增强**：外部工具：调用计算器处理复杂运算，避免数值错误
- **程序辅助**：生成代码（如Python）执行数学计算

[1] Kojima T, Gu S S, Reid M, et al. Large language models are zero-shot reasoners[J]. Advances in neural information processing systems, 2022, 35: 22199-22213.
[2] Guan X, Zhang L L, Liu Y, et al. rStar-Math: Small LLMs Can Master Math Reasoning with Self-Evolved Deep Thinking[J]. arXiv preprint arXiv:2501.04519, 2025.
[3] Wang X, Wei J, Schuurmans D, et al. Self-consistency improves chain of thought reasoning in language models[J]. arXiv preprint arXiv:2203.11171, 2022.

长推理模型

- **长推理模型**：推理时产生长思维链的模型
- **典型代表**：OpenAI-o1[1]，DeepSeek-R1[2]
- **思维链特征**：
  - 自我反思
  - 切换思路
  - 答案验证与纠错
- 通过增长推理时的生成长度，增大了搜索空间，最终提高数学推理的性能
- 思维链中融入了类人的推理策略

[1] https://openai.com/index/introducing-openai-o1-preview/
[2] Guo D, Yang D, Zhang H, et al. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning[J]. arXiv preprint arXiv:2501.12948, 2025.

长推理模型

- 思维链特征:
  - 自我反思
  - 切换思路
  - 答案验证与纠错
- 例子: OpenAI-o1

Moreover, since p is a polynomial of degree $2n$, then $p\left(\frac{1}{x}\right)$ is a rational function of x of degree $2n$ when considered as a function of x.

But we can consider $q(x) = p\left(\frac{1}{x}\right) - x^2$.

Alternatively, let's consider $s(x) = x^{2n}p\left(\frac{1}{x}\right) - x^{2n+2}$. 切换思路

Let me explain: We can consider that $p\left(\frac{1}{x}\right) = x^{-2n} + a_{2n-1}x^{-2n+1} + \ldots + a_0$.

Then multiplying by x^{2n} gives:

Similarly, $m(x) \cdot h(x)$, when evaluated at $x = 0$, becomes $m(0) \cdot h(0) = \left[\prod_{k=1}^{n}(-k^2)\right][-0 + c] = \left[(-1)^n k_1^2 k_2^2 \cdots k_n^2\right] c$.

But since $k = 1$ to n, the product k^2 is

Wait, sorry, actually, $\prod_{k=1}^{n} k^2 = (n!)^2$ 自我纠错

So $m(0) = (-1)^n (n!)^2$.

Therefore, $m(0)h(0) = (-1)^n (n!)^2 (0 + c) = (-1)^n (n!)^2 c$.

Similarly, $s(0) = 1$.

Then $t(x) = x^{2n}p\left(\frac{1}{x}\right) = x^{2n}\left[\left(\frac{1}{x}\right)^{2n} + a_{2n-1}\left(\frac{1}{x}\right)^{2n-1} + \cdots + a_0\right] = 1 + a_{2n-1}x + a_{2n-2}x^2 + \cdots + a_0 x^{2n}$.

So $t(x)$ is a polynomial of degree $2n$.

What about x^{2n+2}? 自我反思

So $s(x) = t(x) - x^{2n+2}$.

Therefore, $s(x)$ is a polynomial of degree $\max(\text{degree of } t(x), \text{degree of } x^{2n+2}) = \text{degree}(x^{2n+2})$.

So $s(x)$ is of degree $2n + 2$.

数学大模型

- 数学大模型：专为数学推理任务优化的大模型
- 典型代表：DeepSeek-Math[1]；Qwen2.5-Math[2]
- 训练集：DeepMath-103K[3]
- 评估数据集：
  - GSM8k[4]: 小学难度的高质量数学数据集，包含8.5k个样本
  - MATH[5]: 高中数学竞赛级别难度数据集，包含12,500个样本
  - AIME: 美国数学邀请赛，每年举办，每年的竞赛包含30道题目，奥赛级别难度

[1] Shao Z, Wang P, Zhu Q, et al. Deepseekmath: Pushing the limits of mathematical reasoning in open language models[J]. arXiv preprint arXiv:2402.03300, 2024.
[2] Yang A, Zhang B, Hui B, et al. Qwen2. 5-math technical report: Toward mathematical expert model via self-improvement[J]. arXiv preprint arXiv:2409.12122, 2024.
[3] He Z, Liang T, Xu J, et al. Deepmath-103k: A large-scale, challenging, decontaminated, and verifiable mathematical dataset for advancing reasoning[J]. arXiv preprint arXiv:2504.11456, 2025.
[4] Cobbe K, Kosaraju V, Bavarian M, et al. Training verifiers to solve math word problems[J]. arXiv preprint arXiv:2110.14168, 2021.
[5] Hendrycks D, Burns C, Kadavath S, et al. Measuring mathematical problem solving with the math dataset[J]. arXiv preprint arXiv:2103.03874, 2021.

DeepSeek-Math：

- **120B数学网页清洗**：迭代过滤Common Crawl数据，避免基准污染
- **代码模型初始化**：基于DeepSeek-Coder-v1.5而非通用LLM，提升符号处理能力
- **GRPO强化学习**：分组相对策略优化，减少显存消耗并提升泛化性，后成为R1模型核心
- **性能**：7B模型达到当时的SOTA

Model	Size	English Benchmarks					Chinese Benchmarks		
		GSM8K	MATH	OCW	SAT	MMLU STEM	CMATH	Gaokao MathCloze	Gaokao MathQA
Closed-Source Base Model									
Minerva	7B	16.2%	14.1%	7.7%	-	35.6%	-	-	-
Minerva	62B	52.4%	27.6%	12.0%	-	53.9%	-	-	-
Minerva	540B	58.8%	33.6%	17.6%	-	63.9%	-	-	-
Open-Source Base Model									
Mistral	7B	40.3%	14.3%	9.2%	71.9%	51.1%	44.9%	5.1%	23.4%
Llemma	7B	37.4%	18.1%	6.3%	59.4%	43.1%	43.4%	11.9%	23.6%
Llemma	34B	54.0%	25.3%	10.3%	71.9%	52.9%	56.1%	11.9%	26.2%
DeepSeekMath-Base	7B	**64.2%**	**36.2%**	**15.4%**	**84.4%**	**56.5%**	**71.7%**	**20.3%**	**35.3%**

Table 2 | Comparisons between DeepSeekMath-Base 7B and strong base models on English and Chinese mathematical benchmarks. Models are evaluated with chain-of-thought prompting. Minerva results are quoted from Lewkowycz et al. (2022a).

数学大模型

- **Qwen2.5-Math:**
  - **预训练语料**：Qwen Math Corpus v2（1T token），含合成数据与中文数学问题
  - **数据生成**：Qwen2-72B-Instruct生成高质量QA对，增强多样性
  - **CoT & TIR融合**：思维链推理+工具集成（如Python解释器），支持双语解析
  - **奖励模型迭代**：Qwen2.5-Math-RM指导SFT数据演化，强化学习阶段应用GRPO
- **结果**：7B模型在MATH基准超越30B-70B开源模型，逼近GPT-4

Model \ Benchmark	En			Zh		
	GSM8K *8-shot*	MATH *4-shot*	MMLU STEM *4-shot*	CMATH *6-shot*	GaoKao Math Cloze *5-shot*	GaoKao Math QA *4-shot*
General Model						
Llama-3.1-8B	56.7	20.3	53.1	51.5	8.5	28.5
Llama-3.1-70B	85.5	41.4	78.1	75.5	11.9	43.3
Llama-3.1-405B	89.0	53.8	-	-	-	-
Qwen2-1.5B	58.5	21.7	44.8	55.6	12.7	35.6
Qwen2-7B	79.9	44.2	67.6	76.7	37.3	51.6
Qwen2-72B	89.5	51.1	79.9	85.4	55.9	72.6
Specific Model						
DeepSeekMath-Base-7B	64.2	36.2	56.5	71.7	20.3	40.7
DeepSeek-Coder-V2-Lite-Base	68.3	38.1	59.5	77.8	25.4	51.3
Internlm2-Math-Base-20B	68.2	30.4	63.0	65.9	16.9	40.2
Qwen2-Math-1.5B	71.3	44.4	50.4	79.6	37.3	50.7
Qwen2-Math-7B	80.4	50.4	65.7	83.2	48.3	57.3
Qwen2-Math-72B	89.1	60.5	79.1	86.4	72.9	69.5
Qwen2.5-Math-1.5B	76.8	49.8	51.3	83.0	47.5	54.1
Qwen2.5-Math-7B	**91.6**	55.4	67.8	85.0	57.6	69.5
Qwen2.5-Math-72B	90.8	**66.8**	**82.8**	**89.7**	**72.9**	**86.3**

数学大模型

DeepMath-103K

- 专为高难度数学任务设计的训练集
- 收集了主流的训练集并进行了大量的数据清理工作，确保和常见数学评测集无重合
- 包含了难度标注以及R1回复

使用DeepMath-103K数据训练的模型能够有效地提高数学能力

Model	MATH 500	AMC 23	Olympiad Bench	Minerva Math	AIME 24	AIME 25	Poly Math
Proprietary Models							
o1-mini	–	–	–	–	63.6	–	–
o3-mini (low effort)	–	–	–	–	60.0	–	–
Zero RL from Base Model							
Qwen-2.5-7B (Team, 2024)	54.8	35.3	27.8	16.2	7.7	5.4	28.1
↳ Open-Reasoner-Zero-7B (Hu et al., 2025)	81.8	58.9	47.9	38.4	15.6	14.4	40.7
↳ Qwen-2.5-7B-SRL-Zoo (Zeng et al., 2025a)	77.0	55.8	41.0	41.2	15.6	8.7	33.1
↳ DeepMath-Zero-7B (Ours)	**85.5**	**64.7**	**51.0**	**45.3**	**20.4**	**17.5**	**42.7**
Qwen-2.5-Math-7B (Team, 2024)	46.9	31.9	15.8	15.5	11.2	4.4	22.7
↳ Qwen-2.5-Math-7B-SRL-Zoo (Hu et al., 2025)	75.8	59.7	37.4	29.9	24.0	10.2	36.0
↳ Qat-Zero-7B (Liu et al., 2025)	80.0	66.7	43.4	40.8	32.7	11.7	40.8
↳ Eurus-2-7B-PRIME (Cui et al., 2025)	80.2	64.7	44.9	42.1	19.0	12.7	38.9
↳ DeepMath-Zero-Math-7B (Ours)	**86.9**	**74.7**	**52.3**	**49.5**	**34.2**	**23.5**	**46.6**
RL from Instruct Models							
R1-Distill-Qwen-1.5B (Guo et al., 2025)	84.7	72.0	53.1	36.6	29.4	24.8	39.9
↳ DeepScaleR-1.5B-Preview (Luo et al., 2025)	89.4	80.3	60.9	42.2	**42.3**	29.6	**46.8**
↳ Still-3-1.5B-Preview (Chen et al., 2025)	86.6	75.8	55.7	38.7	30.8	24.6	43.1
↳ DeepMath-1.5B (Ours)	**89.9**	**82.3**	**61.8**	**42.5**	37.3	**30.8**	46.6
OpenMath-Nemotron-1.5B (Moshkov et al., 2025)	91.8	90.5	70.3	26.3	61.3	50.6	56.8
↳ DeepMath-Omn-1.5B (Ours)	**93.2**	**94.2**	**73.4**	**28.3**	**64.0**	**57.3**	**58.7**

蒸馏学习

- 如何快速获得长推理模型的数学推理能力？
- 监督微调（SFT）
  - 作用：对齐模型输出与人类解题格式，学习标准推理步骤
  - 数据要求：需高质量链式标注
- 知识蒸馏
  - **流程**：大模型（教师）生成CoT数据 → 小模型（学生）SFT学习推理模式
  - **代表方案**：DeepSeek-R1蒸馏版Qwen-7B性能超越32B基线模型

蒸馏学习

- **蒸馏三步法**：
  - **数据生成**：教师模型生成高质量CoT数据
  - **拒绝采样**：筛除错误答案，保留逻辑清晰、格式规范的推理链
  - **SFT微调**：小模型直接学习教师输出，无需额外RL阶段
- **优势**：高效利用数据，门槛较低，能直接学习复杂的推理行为
- **局限**：小模型性能上限依赖教师能力，复杂问题仍需大模型

数学能力评估

- **传统基准**：
  - 数据集： GSM8K（小学）、MATH（高中）、AIME（竞赛）
  - 指标：正确率、Pass@k（采样k次通过率）
- **数学评估特点**：当前研究仅针对有明确数值/符号答案的问题
- **前沿方向**：证明题正确性评估
- **挑战**：
  - LLM自然语言输出：如何从自然语言解答中抽取答案
  - 结果一致性评估：如何评估数学表达式一致性

数学能力评估

- 答案抽取：
  - GSM8K: 使用4个连续的####来标记答案
  - MATH: 使用\\boxed{}来包裹答案，目前主流解决方案
- 优点：
  - 统一抽取标准，方便评估
- 缺点：
  - 模型需要专门针对输出格式训练
  - 无法涵盖所有输出场景

question string · *lengths* 137↔232 46%	answer string · *lengths* 50↔168 19.9%
Natalia sold clips to 48 of her friends in April, and then she sold half as many clips in May. How many clips did Natalia sell altogether in April and May?	Natalia sold 48/2 = <<48/2=24>>24 clips in May. Natalia sold 48+24 = <<48+24=72>>72 clips altogether in April and May. #### 72

GSM8K方案

problem string	solution string
Let \[f(x) = \left\{ \begin{array}{cl} ax+3, &\text{ if }x>2, \\ x-5 &\text{ if } -2 \le x \le 2, \\ 2x-b &\text{ if } x <-2. \end{array} \right.\]Find $a+b$ if the piecewise function is continuous (which means that its graph can be drawn without lifting your pencil from the paper).	For the piecewise function to be continuous, the cases must "meet" at 2 and -2. For example, $ax+3$ and $x-5$ must be equal when $x=2$. This implies $a(2)+3=2-5$, which we solve to get $2a=-6 \Rightarrow a=-3$. Similarly, $x-5$ and $2x-b$ must be equal when $x=-2$. Substituting, we get $-2-5=2(-2)-b$, which implies $b=3$. So $a+b=-3+3=\boxed{0}$.

MATH方案

数学能力评估

- **答案评估**：如何评估模型生成的答案是否正确？
- **规则评估**：
  - 由于模型输出的答案基本遵循LaTex格式，因此可以通过LaTex规则获取表达式，然后通过人工规则判断是否一致
  - 典型代表：Huggingface Math-Verify[1]
- **LLM Judge**：
  - 规则评估无法涵盖文字表达的边缘情况，且答案抽取可能有误
  - 使用大模型（一般经过专门训练）来直接判断生成答案和标准答案是否等价
  - 典型代表： Omni-Judge[2], 一个专用于Omini测试集的答案评估模型

[1] https://github.com/huggingface/Math-Verify
[2] https://huggingface.co/KbsdJames/Omni-Judge

总结

- **数学推理**： LLM通过逻辑分析、符号运算与多步推导解决数学问题的能力，需理解数学概念（如代数、几何）并生成正确答案
- **长推理模型**：推理时产生长思维链的模型
- **数学大模型**：专为数学任务优化，使用数学数据进行预训练的大模型
- **蒸馏学习**：大模型生成数据，小模型直接学习大模型采样结果
- **数学评估**：
  - 常见Benchmark： GSM8K, MATH500, AIME
  - 评估流程：生成结果，通过统一格式抽取结果，使用规则/LLM进行一致性比对

上海交通大學

SHANGHAI JIAO TONG UNIVERSITY

Thank you