【开源版】LinkWise Education CoT Dataset (Chinese High School Mathematics, 10k) 临科智华教育思维链数据集(中国高中数学, 10k)

LinkWise Education CoT Dataset(Chinese High School Mathematics,10k) 临科智华教育思维链数据集(中国高中数学,10k)

随着大语言模型的快速发展,多模态融合、复杂推理能力的提升以及硬件资源的优化利用都取得了显著进展,评分不断攀升,一批新兴 AI 模型不断涌现,并广泛应用于生活的各个领域。然而,在实际使用过程中,模型仍然不可避免地出现幻觉、答案不够流畅甚至错误的情况。这一问题的根源在于,尽管算力和算法不断升级,最关键的数据集质量却未能同步提升,导致模型在复杂多变的真实场景中表现受限。

数学作为逻辑推理和问题求解的核心学科,是衡量大语言模型理解能力和推理深度的重要标准。为了提升 AI 的逻辑水平和思考能力,临科智华率先开源了行业首个针对中国高考数学的数据集: < Education CoT Dataset(Chinese High School Mathematics) > ,该数据集结合思维链(Chain of Thought,以下简称 CoT)的思想,在临科数智引擎(LinkWise Data Engine),以及数百名数据处理工程师共同努力下,通过人机协作与人工验证,构建了一个高质量的数学数据集,能有效提升大模型的数学推理能力及思维能力。

目录 Contents

- 1. 模型训练效果测试
- 2. 构建数据集
- 3. 数据集特点
- 4. 如何帮助大模型提升数学能力
- 5. 数据结构和数据集体量
- 6. 为什么选择数学
- 7. 关于我们
- 8. 使用许可

9. 联系我们

1. 模型训练效果测试 Mode Training Effect Test

1.1 测试结果

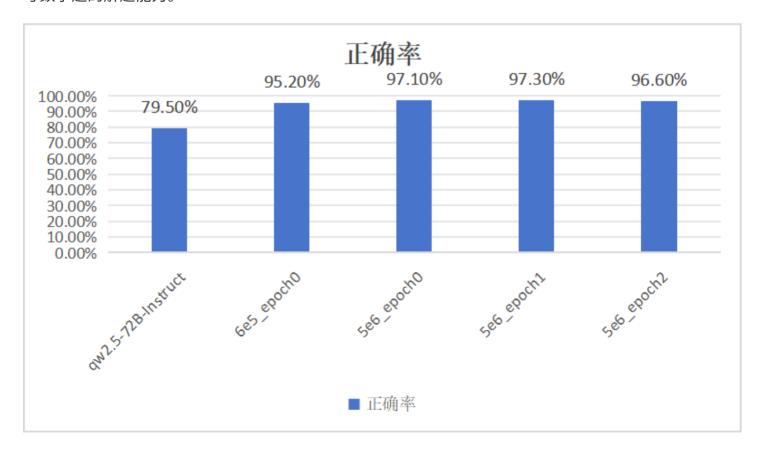
1.1.1 测试结果表格

组别	测试集题 数	正确解答题数	正确率	解答过程平均长度
参照组(基线模型)	1000	795	79.5%	1036
对照组(微调后模型)	1000	973	97.3%	4237

1.1.2 结果分析

1.1.2.1 解答正确率显著提升

微调后的模型在测试集上的正确率由 79.5% 提高至 97.3%,表明该 CoT 数据集能够有效提升模型对高考数学题的解题能力。



1.1.2.2 解答过程的长度增加

参照组的解答过程平均长度为 1036,而微调后的模型解答平均长度增加至 4237。这表明微调后的模型在解题时生成了更详细的推理过程,与 CoT 数据集的训练特性相符合。



1.2 测试方法

为了评估所构建的高中数学 CoT 数据集对模型数学推理能力的提升效果,我们设计了一组对照实验,测试方法如下:

1.2.1 数据集划分

从标注好的 CoT 数据集中随机挑选 9702 道数学题作为训练集、1000 道数学题作为验证集、1000 道数学题作为测试集,确保训练集、验证集和测试集相互之间无交集。

1.2.2 实验设计

1.2.2.1 参照组测试

选择 Qwen2.5-72B-Instruct 作为基线模型,直接使用该模型对测试集进行解答,评估其解答正确率及解答过程的平均长度,作为参照组的测试结果。

1.2.2.2 对照组测试

在基线模型的基础上,使用训练集进行 SFT 全参数微调。训练完成后,使用微调后的模型对测试集进行解答,并评估其解答正确率及解答过程的平均长度,作为对照组的测试结果。

1.2.3 训练环境与参数设定

GPU 环境: 8*H100 80GB GPU

训练框架: DeepSpeed ZeRO-3

主要训练参数: 学习率=5e-6, Epoch=2, Batch Size=64

1.3 结论

综合来看,该实验表明使用高质量的 CoT 数据集进行微调训练能够有效增强大模型在数学推理任务上的表现,同时也使其推理过程更加详尽。

2. 构建数据集 Dataset Construction

构建高质量的数据集并不仅仅是依靠资源和人力的堆砌,否则不过是将试卷从纸上搬到电脑上,无法 真正提升 AI 的推理能力。我们需要的是经过深度处理、能够增强 AI 思维能力的数据。

对此,我们的数据专家给出了答案:利用思维链拆解题目,使 AI 能够更高效地进行推理和计算。思维链并非简单的解题思路,而是一套完整且严密的解题步骤,确保每一步都环环相扣,避免人类思维中的省略和跳步问题,从而引导 AI 正确推演整个问题的解法。此外,数据专家还负责系统流程的整体设计,并持续评估数据集的实际效果,以确保其质量和价值。

目前,主流的标注系统仍然依赖人工进行标注和审核,而我们在这一过程中引入了人工智能进行预标注。AI 能够对采集的数据进行自动整理和逻辑推理,大幅提高标注效率,使得相同时间内的产出量达到过去的 X 倍,而人力则主要负责质量监控,确保数据的精准性。

在数据构建过程中,数据处理工程师负责技术实现与质量把控。我们的数据处理团队均来自高校,其中 85% 为本科学历,5% 为研究生学历,确保了数据的严谨性和答案的准确性,为 AI 的推理能力提升 奠定了坚实基础。

3. 数据集特点

本数据集专为大语言模型(LLMs)在数学推理任务中的训练与评估设计,具备以下技术优势:

- 1. 高准确性: 所有题目来源可靠,均由人工逐题审核,确保表达规范、答案正确、逻辑清晰。即使在复杂题型中,也能保持高标准的数据一致性,适合对大语言模型的基础能力和思维链能力进行训练和评估。
- 结构化标注:每道题目包含题型,难易度,题干,知识点,答案,解题步骤,解题思路,思考过程,适用多种场景。
- 3. 覆盖面广:覆盖高中数学核心模块:代数、函数、几何、数列、概率统计等,题型多样,难度层级分明,兼顾基础题与高阶题,有利于模型全面建模与泛化能力提升。
- 4. AI生成思维链:推理过程由大型语言模型(LLM)生成后,经人工审阅与优化,确保推理链条完整、表达规范。该设计能有效提升模型对逐步推理路径(Step-by-step reasoning)的学习能力,适用于 Chain-of-Thought Prompting、Instruction Tuning 等训练范式。
- 5. 格式标准:数据集采用标准化 JSON,字段命名清晰,兼容 Hugging Face Datasets、Pandas、LangChain 等主流工具,适合快速接入训练/评估管线。

4. 如何帮助大模型提升数学能力 How Does It Work

4.1 原始数据的采集思路

广泛采集各类高中数学真题,模拟题等,涵盖从易到难各级难度,选择题,判断题,填空题,解答题 多类题型,经过严格质检保证题目和答案的准确,原始数据是整个数据集的基础, 保证多样,全面, 准确,才能保证增强数据的质量。

4.2 对原始数据的标注思路

标注环节的重点是思维链,把解题过程拆解成详细步骤,帮助模型建立完整的解题思路,避免人类思维的跳步,省略等问题,提高正确率。

4.3 数据增强的思路

4.3.1 为什么不继续采用原始数据集,而要采用数据增强?

依靠原始数据集,模型虽然可以学习到基础的解题方式,但其思维链和逻辑推理能力难以获得质的提升。而数据增强不仅帮助模型提升在常规题型中的准确率,更通过多样化和复杂化的训练,强化模型的推理深度、逻辑严谨性和自我纠错能力。

4.3.2 数据增强的具体方法

本数据集主要采用以下三种数据增强方法:

题干增强:保持题目其他内容不变,改变叙述方式,提升对文字的理解和逻辑能力。

解题变量增强:保持题目其他内容不变,更换变量,提升推理能力。

数字增强:保持题目其他内容不变,改变数字,提升计算能力。

4.3.3 数据增强示例

【原题】

已知函数 $f(x)=egin{cases} x^2,0\leq x\leq 1\ |\ln(x-1)|,x>1 \end{cases}$,若方程 f(x)=kx-2 有两个不相等的实数根,则实数 k的取值可以是()。

A.
$$\frac{1}{2}$$
 B. 1 C. 2 D. 3

答案: D

【经过题干增强的新题目】

设函数 f(x) 在区间 [0,1] 上取值为 x^2 ,而在 x>1 时取值为 $|\ln(x-1)|$ 。若直线 y=kx-2 与函数 y=f(x) 的图像存在两个不同的交点,则实数 k 的可能取值是()。

A.
$$\frac{1}{2}$$
 B. 1 C. 2 D. 3

答案: D

【经过数字增强的新题目】

已知函数 $f(x)=egin{cases} 2x^2, 0\leq x\leq 1 \\ |\ln(4x-3)|, x>1 \end{cases}$,若方程 f(x)=kx-5 有两个不相等的实数根,则实数 k 的取值可以是()。

A. 6 B. 8 C. 10 D. 12

答案:B

【经过解题变量增强的新题目】

已知函数
$$g(t)=egin{cases} t^2, & 0\leq t\leq 1 \ \ln(t-1)|, & t>1 \end{cases}$$
 ,

若直线 y=kt-2与函数 y=g(t) 的图像恰好有两个不同的交点,则实数 k 的可能取值是()。

A.
$$\frac{1}{2}$$
 B.1 C.2 D.3

答案: D

使用经过数据增强的数据集进行训练,可以进一步提高模型的泛化能力,减少过拟合,提高在实际应 用中的表现。

5. 数据结构和数据集体量 Data Structure and Data Volumn

本开源数据集由训练集和测试集组成,训练集共9375道题,测试集共977道题。

5.1 训练集构成

题型	百分比	题目数量
选择题	66.56%	6244
填空题	18.05%	1693
解答题	15.38%	1438

题目难易度	百分比	题目数量
简单	15.40%	1438
中等	74.42%	6983
困难	10.19%	954

知识点	百分比	题目数量
函数与导数	43.25%	4054
解析几何	18.63%	1746
不等式	17.20%	1612
三角函数与 解三角形	14.94%	1400
数列	10.03%	940
集合与常用逻辑用语	9.81%	919
平面向量	8.98%	841
概率与统计	7.62%	714
排列组合与	7.08%	663
复数	6.75%	632
立体几何与 空间向量	4.81%	450

5.2 测试集构成

题型	百分比	题目数量
选择题	91.82%	909
填空题	8.18%	81

题目难易度	百分比	题目数量
简单	23.33%	231
中等	68.38%	677
困难	7.37%	73

知识点	百分比	题目数量

函数与导数	41.41%	410
解析几何	12.32%	122
不等式	16.77%	166
三角函数与解三角形	12.42%	123
数列	9.09%	90
集合与常用 逻辑用语	13.74%	136
平面向量	12.42%	67
概率与统计	5.05%	50
排列组合与 二项式定理	7.37%	73
复数	9.19%	91
立体几何与 空间向量	4.95%	49

5.3 数据结构示例

每道试题包含:

- 1 {"job_id": 1000031602,
- 2 "volume_name": "北京市通州区2022届高三查漏补缺练习数学试题",
- 3 "question_type": "填空题",
- 4 "is_required": "是",
- 5 "difficulty": "中等",
- 6 "question_stem": "直线\$x\\cos\\theta+\\sqrt{3}y + 2 = 0\$的倾斜角的范围是_____。",
- 7 "knowledge_points": "1.直线的一般式方程与斜截式方程的转化。\n\n2.斜率与倾斜角的关系。\n\n3.三角函数的有界性。\n\n4.反三角函数在求解倾斜角中的应用。",
- 8 "answer": "\$\\left[0, \\frac{\\pi}{6}\\right] \\cup \\left[\\frac{5\\pi}{6},
 \\pi\\right)\$",

- 10 "solution": "首先将直线方程\$x\\cos\\theta+\\sqrt{3}y + 2 = 0\$转化为斜截式\$y = kx + b\$,得到斜率\$k = -\\frac{\\cos\\theta}{\\sqrt{3}}\$。根据余弦函数的取值范围\$\\cos\\theta \\in [-1,1]\$,可得斜率\$k\$的范围为\$[-\\frac{1}{\\sqrt{3}}},\\frac{1}{\\sqrt{3}}]\$。再通过斜率与倾斜角的关系\$\\tan\\alpha = k\$,结合正切函数在不同象限的单调性,最终确定倾斜角的范围。",
- "thinking": "嗯,我现在要解这个题目: 直线 $x\cos\theta + \sqrt{3} y + 2 = \theta$ 的倾斜角的范围是多少。首 11 先,我得回忆一下直线方程的一般形式以及倾斜角相关的知识。好的,直线的倾斜角是指直线与x轴正 方向之间最小的正角,范围在0到 π 之间。而斜率k就是 $tan\alpha$,其中 α 是倾斜角。所以可能要先把这个直 线方程转化为斜截式,找到斜率,再根据斜率的范围来确定倾斜角的范围。\n\n首先,原方程是 $x\cos\theta$ + √3 y +2 =0。我需要把它整理成y = kx + b的形式。移项的话,√3 y = $-x\cos\theta$ -2, 所以两边除以 $\sqrt{3}$ (− cos θ / $\sqrt{3}$) x − 2/ $\sqrt{3}$ 。那这里的斜率k就是 −cos θ / $\sqrt{3}$ 。接下来,我得分析 这个斜率的取值范围,然后根据 $tan\alpha = k$ 来确定 α 的范围。 $\n\n$ 接下来, $cos\theta$ 的范围呢?因为 θ 是任 意实数,所以 $\cos\theta$ 的范围是[-1,1]。所以 $-\cos\theta$ 的范围其实是[-1,1]取反,也就是[-1,1]但是乘 以-1,即[-1,1]的范围变成从-1到1,对吗?不,等一下,cosθ∈[-1,1],所以当乘以-1的时候,- $\cos\theta$ ∈ [-1,1],即当 $\cos\theta$ 取1时, $-\cos\theta$ =-1;当 $\cos\theta$ 取-1时, $-\cos\theta$ =1。所以 $-\cos\theta$ 的范围确实是 [-1,1]吗?或者说其实是[-1,1]中的每个值都可能被取到?是的,没错,因为当 θ 取不同的值时, cosθ可以覆盖整个[-1,1],所以-cosθ的范围同样是[-1,1]。\n\n所以斜率k = -cosθ / √3,所 以k的取值范围就是 $[-1/\sqrt{3}, 1/\sqrt{3}]$ 。因为当 $-\cos\theta$ 最大为1时, $k=1/\sqrt{3}$;当 $-\cos\theta$ 最小为-1时, k=-1/√3。所以k的范围是[-1/√3, 1/√3]。nn接下来,倾斜角α的范围要根据k的范围来确定。已 知k = tanα,所以需要找出所有可能的α \in [0, π)使得tanα \in [-1/ $\sqrt{3}$, 1/ $\sqrt{3}$]。这个时候,我要考 虑tanα在不同区间的取值情况。\n\n首先,tanα在 $(-\pi/2, \pi/2)$ 的基本周期内,tanα= $1/\sqrt{3}$ 对应的α 是 π /6,而tanα=-1/ $\sqrt{3}$ 对应的α是- π /6,但倾斜角的范围是[0, π)。所以当tanα在[-1/ $\sqrt{3}$,1/ $\sqrt{3}$]之 间时,对应的 α 应该分为两部分考虑:当k≥0时,即 $tan\alpha$ ∈ [0,1/√3],对应的 α ∈ $[0,\pi/6]$;而当k≤0时, $tan\alpha \in [-1/\sqrt{3},0]$,对应的 $\alpha \in [5\pi/6,\pi)$,因为在第二象限, $tan\alpha$ 为负,但倾斜角α是取那个角 度的补角,也就是π加上这个负角的正切对应的角度。不过这里需要注意,倾斜角α的范围是θ到π,所 以当k为负数时,对应的 α 应该是在 $\pi/2$ 到 π 之间的角度,具体来说, $\tan \alpha$ 为负数时, α 在 ($\pi/2$, π)之 间,此时 $tan\alpha = tan(\pi - \beta)$,其中β是一个锐角。例如, $tan\alpha = -1/\sqrt{3}$ 时,对应的 α 是 $5\pi/6$,因为 $tan(5\pi/6) = tan(\pi - \pi/6) = -tan(\pi/6) = -1/\sqrt{3}$ 。\n\n综上,当k \in [-1/ $\sqrt{3}$, 1/ $\sqrt{3}$]时,倾 斜角 α 的范围应该是 $[0, \pi/6] \cup [5\pi/6, \pi)$ 。不过题目问的是倾斜角的范围,所以需要把这些区间 的闭包写出来。这种情况下,答案应该是 $\alpha \in [0, \pi/6] \cup [5\pi/6, \pi)$,对吗? \n\n不过还要再仔细 检查一下。当k的范围是[-1/√3, 1/√3],这时候斜率绝对值不超过1/√3。当斜率为正时,最大到1/√3 $\sqrt{3}$,对应的角度是 $\pi/6$;当斜率为负时,最小到 $-1/\sqrt{3}$,对应的角度是 $5\pi/6$ 。这时候倾斜角的范围确 实是0到 π /6,以及5 π /6到 π 之间。对吗?\n\n是的,没错。因为当斜率k在[$-1/\sqrt{3}$, $1/\sqrt{3}$]之间时, α 在0到 π /6或者 5π /6到 π 之间。所以答案应该是这两个区间的并集。所以最终答案应该写成区间形式, 或者用不等式,比如0 $\leq \alpha \leq \pi/6$ 或 $5\pi/6 \leq \alpha < \pi$ 。不过一般来说,题目中可能要求写成闭区 间,但因为π本身对应的是水平向左的情况,这里可能当斜率为0时,倾斜角是0,而当斜率为负的时 候,最大到π,但实际上当斜率为负且趋近于0时,倾斜角趋近于π。不过严格来说,倾斜角是取在

 $[0,\pi)$ 的,所以当k=0时,倾斜角是0,而当k趋近于0负方向时,倾斜角趋近于 π 。所以这里当 k \in $[-1/\sqrt{3},1/\sqrt{3}]$ 时,对应的倾斜角范围是两个闭区间,即 $[0,\pi/6]$ 和 $[5\pi/6,\pi)$ 。但是,当k=1/ $\sqrt{3}$ 时,对应 $\alpha=\pi/6$;当k= $-1/\sqrt{3}$ 时,对应 $\alpha=5\pi/6$ 。而斜率为0时对应 $\alpha=0$ 。不过题目中的直线方程是存在的,当cos $\theta=0$ 的时候,k=0吗?因为原来的方程xcos $\theta+\sqrt{3}$ y +2 = 0,当cos $\theta=0$ 时,方程变为 $\sqrt{3}$ y +2=0,即y= $-2/\sqrt{3}$,这是一条水平线,所以倾斜角是0。所以这里的k=0确实会被包含进去。 $\ln \pi$ 中价,倾斜角的范围是0到 $\pi/6$ 以及 $\pi/6$ 到 π 。所以最终答案应该是这两个区间的并集。",

12 "remark": null}

字段	解释
job_id	每道题的唯一标识符
volume_name	试卷名
question_type	题型
is_required	是否必做
difficulty	难易度
question_stem	题干
knowledge_points	知识点
answer	答案
reasoning	解题步骤
solution	解题思路
thinking	思考过程
remark	备注

5.4 数据增强示例

原题:

```
8
    B. 31
9
    C. 33
10
11
   D. 36
          ш,
12
           "知识点":「
13
14
               "等比数列的定义及通项公式。",
               "等差中项的定义。",
15
              "等比数列前n项和公式。"
16
17
           ],
           "解题思路":「
18
              "(1)根据等比数列的性质,以及等差中项的定义,列出关于首项 $a_1$ 和公比
19
    $q$ 的方程组,通过解方程组求出 $a 1$ 和 $q$,最后利用等比数列前 $n$ 项和公式计算
    $S_5$ 的值。"
20
           ],
           "解题步骤": [
21
               "(1) 设此正项等比数列的首项为 $a_1$, 公比为 $g$ ($g > 0$)。
22
23
    由 $a 2 \cdot a 5 = 2a 3$ 可得:
24
    $a_2 \cdot a_5 = 2a_3$
    根据等比数列通项公式:
25
26
    $$a 1 q \cdot dot a 1 q^4 = 2a 1 q^2$$
   化简得:
27
    $$a 1^2 q^5 = 2a 1 q^2$
28
29
    两边同除 $a_1 q^2$ (因 $a_1 > 0$, $q > 0$):
    $a_1 q^3 = 2 \tan{1}$
30
31
    (2) 由 $a_4$ 与 $a_6$ 的等差中项为 $\frac{5}{4}$ 得:
32
    \frac{a_4 + a_6}{2} = \frac{5}{4} \le a_4 + a_6 = \frac{5}{2}
33
    代入通项公式:
34
    $$a_1 q^3 + a_1 q^5 = \frac{5}{2}$
35
    将 $(1)$ 式 $a_1 q^3 = 2$ 代入:
36
    $$2 + 2g^2 = \frac{5}{2}$$
37
    解得:
38
    $$2q^2 = \frac{1}{2} \leq q^2 = \frac{1}{4}$
39
40
    因 $q > 0$,故:
41
    $$q = \frac{1}{2}  \tag{2}$$
42
    (3) 将 $(2)$ 代入 $(1)$ 求 $a_1$:
43
    $a_1 \cdot \left( \frac{1}{2} \right)^3 = 2 \implies a_1 \cdot \frac{1}{8} =
44
    2$$
    解得:
45
    $$a_1 = 16$$
46
47
    (4) 计算前 $5$ 项和 $S_5$:
48
    $$S_5 = \frac{1 - q^5}{1 - q} = \frac{16 \left(1 - \left(1\right)}{2}
49
    \  \in \{1 - frac\{1\}\{2\}\} 
    化简得:
50
```

```
51
            $$S_5 = \frac{1}{2} = 32 \cdot \frac{31}{32}}{\frac{1}{2}} = 32 \cdot \frac{31}{32} = 32 \cdot \frac{31}
            31$$ ",
52
                                 ],
                                 "思考过程":{
53
                                 "嗯,我现在要解决这个关于正项等比数列的问题,题目是已知a,a,=2a,,而且a。和a,的等差
54
            中项是5/4, 求S<sub>5</sub>的值。选项是A到D, 我需要一步步来。
55
            首先,我应该回忆一下等比数列的基本性质。等比数列的通项公式是a = a_1q^n(n-1),其中a_1是首
56
            项, g是公比。这里题目说是正项等比数列, 所以g应该大于0, 而且所有的a 都是正数。
57
            题目给出的第一个条件是a,a,=2a,。根据通项公式,a,是a,q,a,是a,q^4,所以它们的乘积就是a,q *
58
            a_1g^4 = a_1^2g^5。而右边的2a_1则是2*a_1g^2。于是方程应该是a_1^2g^5 = 2a_1g^2。接下来可能需要化简
            这个式子,两边都除以a,q²(因为a,和q都是正数,所以可以这么做),得到a,q³=2。这一步应该没问
            题,对吧?
59
            接下来第二个条件,a_a和a_b的等差中项是5/4。等差中项的定义是两个数的平均数,所以(a_a+
60
            a。)/2 =5/4,也就是a。+ a。=5/2。同样用通项公式代入,a。是a,q³,a。是a,q^5,所以方程变成
            a_{1}q^{3} + a_{1}q^{5} = 5/2。这时候注意到前面第一个条件已经得到了a_{1}q^{3} = 2,所以可以把a_{1}q^{3}代入进去,
            得到2 + 2q^2=5/2。因为a_1q^3=2,所以a_1q^5=a_1q^3*q^2=2q^2。这样代入的话,整个式子就是2 +
            2q<sup>2</sup>=5/2,解得2q<sup>2</sup>=5/2 - 2=1/2,所以q<sup>2</sup>=1/4。由于q是正数,所以q=1/2。
61
            接下来,求a,。根据a,q³=2,已知g=1/2,所以g³=1/8,代入得a,*(1/8)=2,所以a,=2*8=16。
62
63
            现在有了a_1=16,g=1/2,那么求S_5,即前五项的和。等比数列前n项和公式是S_5=a_1(1-q^n)/(1-q^n)
64
            q)。代入数值得到S₅=16*(1 - (1/2)^5)/(1 - 1/2)。计算一下,(1/2)^5=1/32,所以1 -
            1/32=31/32, 分母1 - 1/2=1/2, 所以整个式子就是16*(31/32)/(1/2)=16*(31/32)*2=16*
            (31/16)=31。所以答案应该是选项B, 31。
65
            不过,我需要再检查一下每一步是否正确,避免计算错误。比如在第二个条件中,代入是否正确。
66
            a_1q^3=2,那么a_1q^5=2*q^2,对吗? 是的,因为a_1q^5=a_1q^3*q^2=2*q^2。然后方程变为2 +
            2q²=5/2,解这个方程确实得到q²=1/4,q=1/2。接着求a,,没问题。最后计算S₅的时候,分母是1
            - 1/2=1/2,分子是31/32,所以16*(31/32)/(1/2)=16*(31/32)*2=16*(31/16)=31,没错。所
            以答案是B选项,31。"
67
                                 }
68
                                 "答案": {
                                           "B"
69
70
                                              }
71
                    }
72 }
```

题干增强:

```
1 json{
2 {
3 "难易程度": "中等",
```

```
"题干": "已知正项等比数列 $\{a n\}$ 的前 $n$ 项和为 $S n$, 若 $a 2$ 与
    $a 5$ 的乘积等于 $2$ 倍的 $a 3$, 且 $a 4$ 与 $a 6$ 的等差中项为 $\frac{5}{4}$,则
    $S_5$ 的值为()。
5
    A. 29
6
7
    B. 31
8
9
10
    C. 33
11
    D. 36
12
           "知识点": [
13
               "等比数列的定义及通项公式。",
14
               "等差中项的定义。",
15
               "等比数列前n项和公式。"
16
17
           ],
           "解题思路": [
18
               "(1) 根据等比数列的性质,以及等差中项的定义,列出关于首项 $a 1$ 和公比
19
    $q$ 的方程组,通过解方程组求出 $a 1$ 和 $q$,最后利用等比数列前 $n$ 项和公式计算
    $S_5$ 的值。"
20
           ],
           "解题步骤": [
21
               "(1)设正项等比数列的首项为 $a_1$,公比为 $q$ ($q > 0$)。
22
23
    由 $a 2 \cdot a 5 = 2a 3$ 得:
24
    $a_1 q \cdot d = 2a_1 q^2
    化简为:
25
    $a_1^2 q^5 = 2a_1 q^2$
26
    两边同除 $a 1 g^2$ (因 $a 1 > 0$, $g > 0$):
27
    $a_1 q^3 = 2 \tan{1}$
28
29
    (2) 由 $a_4$ 与 $a_6$ 的等差中项为 $\frac{5}{4}$ 得:
30
    $$a 4 + a 6 = \frac{5}{2}$
31
32
    代入通项公式:
    $a_1 q^3 + a_1 q^5 = \frac{5}{2}$
33
34
    将 $(1)$ 式 $a_1 q^3 = 2$ 代入:
35
    $$2 + 2q^2 = \frac{5}{2}$
    解得:
36
    q^2 = \frac{1}{4} \le q = \frac{1}{2} 
37
38
    (3) 将 $q = \frac{1}{2}$ 代入 $(1)$ 得:
39
    \frac{1}{2} \cdot \frac{1}{2} \cdot 3 = 2 \cdot \frac{1}{2} \cdot 3 = 2 \cdot \frac{1}{2} \cdot 3 = 16
40
41
    (4) 计算前 $5$ 项和 $S 5$:
42
    $$_5 = \frac{16 \left( 1 - \left( \frac{1}{2} \right)^5 \right)}{1 - \frac{1}{2}}
43
    {2}} = \frac{16 \cdot \frac{31}{32}}{\frac{1}{2}} = 31$$ "
44
           ],
           "思考过程":{
45
```

```
"好的,我现在要解决这个关于正项等比数列的问题。题目是说已知正项等比数列{a n}的
46
    前n项和为S n,满足a,和a。的乘积等于2倍的a,,而且a。和a。的等差中项是5/4,求S。的值,选项是A到
    D中的B选项31。我需要一步步来解。
47
    首先,先理清楚已知条件。题目给出的是一个正项等比数列,也就是说公比q是正数,首项a<sub>1</sub>也是正
48
    数。接下来有两个条件:
49
    1. a_2 \cdot a_5 = 2a_3
50
    2. a_a和a_s的等差中项是5/4,也就是(a_a + a_s)/2 = 5/4,即a_a + a_s = 5/2
51
52
    目标是求前五项和S。的值。选项给出的是B选项31,所以可能需要验证这个结果是否正确,但先不管选
53
    项,自己推导一遍。
54
55
    首先,等比数列的通项公式是a_n = a_i \cdot q^{n-1},所以各个项都可以用a_i和q来表示。
56
    首先处理第一个条件: a,·a, = 2a,
57
58
59
    代入通项公式:
60
61
    a_2 = a_1 \cdot q^{2-1} = a_1q
62
63
    a_5 = a_1 \cdot q^{5-1} = a_1 q^4
64
65
    a_3 = a_1 \cdot q^{1} = a_1 q^{2}
66
    所以左边是a₂·a₅ = (a₁q)(a₁q⁴) = a₁²q⁵
67
68
    右边是2a, = 2·a,q<sup>2</sup>
69
70
    因此方程变为: a,²q⁵ = 2a,q²
71
72
73
    两边都是正数,可以两边同时除以a,q^2,得到a,q^3 = 2。这应该是一个方程,记为方程(1)。
74
75
    接下来处理第二个条件: a<sub>4</sub> + a<sub>6</sub> = 5/2
76
    同样代入通项公式:
77
78
79
    a_4 = a_1 q^3
80
81
    a_6 = a_1q^5
82
    所以方程是a,q³ + a,q⁵ = 5/2
83
84
    这里可以提取公因子a_1q^3,得到a_1q^3(1 + q^2) = 5/2
85
```

而根据方程(1), $a_1q^3 = 2$,所以代入进去,得到 $2(1 + q^2) = 5/2$

```
解这个方程:
 89
 90
     两边乘以2,得到4(1 + q^2) = 5 \Rightarrow 4 + 4q^2 = 5 \Rightarrow 4q^2 = 1 \Rightarrow q^2 = 1/4 \Rightarrow q = 1/2 (因
 91
     为q是正数)
 92
     所以g = 1/2,代入方程(1)求a,:
 93
 94
 95
     a_1*(1/2)^3 = 2 \Rightarrow a_1*(1/8) = 2 \Rightarrow a_1 = 2 * 8 = 16
96
     现在有了a,=16,q=1/2,接下来求前五项的和S。
 97
98
     等比数列前n项和的公式是S_n = a_1(1 - q^n)/(1 - q)
99
100
     代入数值:
101
102
     S_s = 16*(1 - (1/2)^5)/(1 - 1/2) = 16*(1 - 1/32)/(1/2) = 16*(31/32)/(1/2) = 16*
103
     (31/32)*2 = 16*(31/16) = 31
104
105
     所以S<sub>5</sub>的值是31,对应选项B。
106
     不过,我需要检查一下是否有哪里算错。首先,在第一个条件推导的时候,确实得到a,q³=2。第二个
107
     条件代入之后得到q=1/2,然后计算a,确实得到16,代入前五项和的计算,步骤正确,结果确实是
     31, 所以答案应该是B选项。"
108
             "答案": {
109
                 "B"
110
111
                  }
112
        }
113
     }
```

数字增强:

```
1
    json{
2
    {
           "难易程度": "中等",
3
           "题干": "已知正项等比数列{$b n$}的前n项和为$T {n}$, 若$b {2} \cdot b {5}
4
    = 4 b_{3}$, 且$b_{4}$与$b_{6}$的等差中项为 $\frac{5}{2}$, 则$T_{5} =$()。
5
    A. 58
6
7
8
    B. 60
9
    C. 62
10
11
    D. 64",
12
```

```
"知识点":「
13
               "等比数列的定义及通项公式。",
14
              "等差中项的定义。",
15
              "等比数列前n项和公式。"
16
17
           ٦,
           "解题思路":[
18
               "(1) 根据等比数列的性质,以及等差中项的定义,列出关于首项 $b 1$ 和公比
19
    $q$ 的方程组,通过解方程组求出 $b 1$ 和 $q$,最后利用等比数列前 $n$ 项和公式计算
    $T 5$ 的值。"
20
           ],
           "解题步骤":「
21
              "(1)设正项等比数列 $\{b n\}$ 的首项为 $b 1$,公比为 $q$ ($q > 0$)。
22
23
    由 $b 2 \cdot b 5 = 4b 3$ 得:
    $$b_1 q \cdot dot b_1 q^4 = 4b_1 q^2$$
24
25
    化简为:
    $$b_1^2 q^5 = 4b_1 q^2$
26
27
    两边同除 $b_1 q^2$ (因 $b_1 > 0$, $q > 0$):
    $$b_1 q^3 = 4 \tan{1}$
28
29
    (2) 由 $b 4$ 与 $b 6$ 的等差中项为 $\frac{5}{2}$ 得:
30
    $$b 4 + b 6 = 5$$
31
   代入通项公式:
32
33
    $$b_1 q^3 + b_1 q^5 = 5$
    将 $(1)$ 式 $b_1 q^3 = 4$ 代入:
34
    $$4 + 4q^2 = 5$$
35
    解得:
36
37
    \$$q^2 = \frac{1}{4} \leq q = \frac{1}{2} \frac{2}{\$}
38
    (3) 将 $g = \frac{1}{2}$ 代入 $(1)$ 得:
39
40
    $b_1 \cdot \left( \frac{1}{2} \right)^3 = 4 \implies b_1 \cdot \frac{1}{8} =
    4 \implies b_1 = 32$$
41
42
    (4) 计算前 $5$ 项和 $T_5$:
43
    $T_5 = \frac{1}{1 - q^5}{1 - q} = \frac{32 \left(1 - \left(\frac{1}{2}\right)}{1 - q^5}\right)}
    \left(1}{2}\right)^5 \right] - \left(1}{2}\right] = \frac{32}{32}{\frac{1}{32}}{\frac{1}{32}}
    \{2\}\} = 62\$\$ "
44
           1,
           "思考过程":{
45
           "嗯,用户现在给了一个新的等比数列问题,不过这次数列是{b_n},前n项和是T_n。题目
46
    条件和之前的类似,但数值有变化。我需要仔细看看这些变化,确保解题过程中不会出错。
47
    首先,题目条件有两个:第一,b₂乘以b₅等于4倍的b₃;第二,b₄和b₅的等差中项是5/2。需要求T₅的
48
    值。选项是A到D,数值分别是58、60、62、64。
49
```

50 好的,先回忆一下之前的问题。在之前的例子中,用户给出的数列是正项等比数列{a_n},同样有类似的条件,通过设立方程解出公比q和首项a,,然后计算前五项的和S。现在的问题应该是类似的,只

```
是参数变了,比如原来的等式是a,a,=2a,,现在变成了b,b,=4b,;原来的等差中项是5/4,现在是
5/2。可能用户希望我按照同样的方法来解决这个问题,但需要调整数值,并注意可能的计算差异。
首先,按照惯例,设正项等比数列{b_n}的首项为b,,公比为q,且q>0。因为是正项数列,所以所有
的项都是正的,不会有负数或者零的情况。
根据第一个条件,b_1 \cdot b_2 = 4b_3。等比数列的通项是b_1 = b_1 \cdot q^n(n-1)。因此:
b_2 = b_1 \cdot q^{(2-1)} = b_1 q
b_5 = b_1 \cdot q^{(5-1)} = b_1 q^4
b_3 = b_1 \cdot q^{(3-1)} = b_1 q^2
所以代入条件得:
b_2 \cdot b_5 = 4b_3 \rightarrow (b_1q) \cdot (b_1q^4) = 4(b_1q^2)
左边是b_1^2 q^5,右边是4b_1 q^2。接下来整理方程:
b_1^2 q^5 = 4b_1 q^2
两边同时除以b_1 q^2(因为b_1>0,q>0,所以可以除),得到:
b<sub>1</sub> q^3 = 4 → 方程(1)
接下来处理第二个条件,即b<sub>4</sub>和b<sub>5</sub>的等差中项是5/2。等差中项的定义是两数之和的一半等于该中项,
因此:
(b_4 + b_6)/2 = 5/2 \rightarrow b_4 + b_6 = 5
同样用通项公式代入:
b_4 = b_1 q^3
b_6 = b_1 q^5
所以:
```

83 84

85

8687

88 89

90 91

9293

 $b_1 q^3 + b_1 q^5 = 5$

现在知道了q=1/2,代入方程(1)求b,:

现在,方程(1)已经给出b, q³ = 4, 所以可以将这个代入上式:

 $4 + 4 q^2 = 5 \rightarrow 4 q^2 = 5 - 4 \rightarrow 4 q^2 = 1 \rightarrow q^2 = 1/4 \rightarrow q = 1/2$ (因为q > 0)

51

52

5354

55 56

57 58

59

60 61

6263

646566

67 68

69

70 71

72 73

74

75 76

77

78 79

80 81 82

```
94
      b_1*(1/2)^3 = 4 \rightarrow b_1*(1/8) = 4 \rightarrow b_1=4*8=32
 95
      接下来求前五项的和T<sub>4</sub>,等比数列前n项和的公式是:
 96
 97
      T = b_1(1 - q^n)/(1 - q)
 98
 99
100
      代入n=5, b<sub>1</sub>=32, q=1/2:
101
102
      T_s = 32*(1 - (1/2)^5)/(1 - 1/2) = 32*(1 - 1/32)/(1/2) = 32*(31/32)/(1/2) = 32*
      (31/32)*2 = 31*2 = 62
103
      所以选项中C选项是62,所以答案应该是C。不过仔细检查计算步骤是否正确。
104
105
106
      再次核对:
107
      1. 由b_2 \cdot b_5 = 4b_3得b_1^2 q^5 = 4b_1 q^2 \rightarrow b_1 q^3 = 4。
108
109
      2. 由等差中项得b₄ +b₆=5 → bˌ q³ +bˌ q⁵=5。代入bˌ q³=4,得到4 +4 q²=5 → q²=1/4
110
      \rightarrow q=1/2_{\circ}
111
      3. 代入得b,=4/(q³)=4/(1/8)=32。
112
113
114
      4. T_s=32*(1 - (1/2)^5)/(1 -1/2)=32*(31/32)/(1/2)=31/(1/2)=62
115
      是的,计算是正确的,所以答案应该是C选项62。"
116
117
               }
               "答案": {
118
                   "C"
119
120
                    }
         }
121
      }
122
```

变量增强:

```
json{
1
2
    {
          "难易程度": "中等",
3
          "题干": "已知正项等比数列 ${b_n}$ 的前 $n$ 项和为$T_{n}$,若 $b_{2} \cdot
4
   b_{5} = 2b_{3}$,且 $b_{4}$ 与 $b_{6}$ 的等差中项为 $\frac{5}{4}$,则 $T_{5} =$
   ()。
5
   A. 29
6
7
8
   B. 31
9
```

```
10
                C. 33
11
12
                 D. 36",
                                               "知识点": [
13
                                                              "等比数列的定义及通项公式。",
14
                                                              "等差中项的定义。",
15
                                                              "等比数列前n项和公式。"
16
17
                                               ],
                                               "解题思路":「
18
                                                              "(1)根据等比数列的性质,以及等差中项的定义,列出关于首项 $b 1$ 和公比
19
                  $q$ 的方程组,通过解方程组求出 $b 1$ 和 $q$,最后利用等比数列前 $n$ 项和公式计算
                  $T_5$ 的值。"
                                               ٦,
20
                                               "解题步骤": [
21
                                                              "(1)设正项等比数列 $\{b n\}$ 的首项为 $b 1$,公比为 $q$ ($q > 0$)。
22
                  由 $b 2 \cdot b 5 = 2b 3$ 得:
23
24
                  $$b_1 q \cdot dot b_1 q^4 = 2b_1 q^2$$
25
                 化简为:
                  $$b_1^2 q^5 = 2b_1 q^2$
26
                  两边同除 $b_1 q^2$ (因 $b_1 > 0$, $q > 0$):
27
                  $$b_1 q^3 = 2 \text{1}}$
28
29
                  (2) 由 $b 4$ 与 $b 6$ 的等差中项为 $\frac{5}{4}$ 得:
30
                  $$b_4 + b_6 = \frac{5}{2}$
31
                代入通项公式:
32
                 $$b_1 q^3 + b_1 q^5 = \frac{5}{2}$
33
                 将 $(1)$ 式 $b_1 q^3 = 2$ 代入:
34
                 $$2 + 2q^2 = \frac{5}{2}$$
35
                 解得:
36
37
                 2^2 = \frac{1}{2} \le q^2 = \frac{1}{4}
                 因 $q > 0$, 故:
38
                 $$q = \frac{1}{2}  \tag{2}$
39
40
41
                 (3) 将 $q = \frac{1}{2}$ 代入 $(1)$ 得:
42
                  $$b_1 \cdot \left( \frac{1}{2} \right)^3 = 2 \in b_1 = 16$
43
                 (4) 计算前 $5$ 项和 $T_5$:
44
                  f = \frac{1}{2} 
45
                  \left(1\right)^5 \right] - \left(1\right)^5 - \left
                 {2}} = 31$$ "
                                               ],
46
                                               "思考过程": {
47
                                               "好,我现在来仔细看看这个等比数列的题目。题目是说有一个正项等比数列{b n},前n项
48
                 和是T_n。已知b,乘以b,等于2倍的b,,同时b,和b,的等差中项是5/4,求T,的值,选项是A到D。
49
```

```
首先,我需要回忆一下等比数列的基本性质。等比数列的通项公式是b_n = b_1*q^n(n-1),其中b_1是首
50
    项,q是公比,而且这里题目说是正项等比数列,所以q应该大于0。
51
    题目给出的第一个条件是b,·b, = 2b,。那我先写出这些项的具体表达式:
52
53
    b_2 = b_1*q^(2-1) = b_1*q
54
55
    b_5 = b_1*q^{(5-1)} = b_1*q^4
56
57
58
    b_3 = b_1 * q^{(3-1)} = b_1 * q^2
59
    所以,代入第一个条件:
60
61
62
    b_2 \cdot b_5 = (b_1 * q) * (b_1 * q^4) = b_1^2 * q^5
63
    右边是2b<sub>3</sub>,即2*b<sub>1</sub>*q<sup>2</sup>
64
65
    因此,方程就是:
66
67
    b_1^2 * q^5 = 2 * b_1 * q^2
68
69
    接下来,我可以两边都除以b,*q²,因为b,和q都是正数,不会为零。这样得到:
70
71
    b_1*q^3 = 2 ... (1)
72
73
    接下来第二个条件是b。和b。的等差中项是5/4。等差中项的定义是两个数的平均数,所以:
74
75
    (b_4 + b_6)/2 = 5/4
76
77
    因此, b_a + b_b = 5/2 * 2 = 5/2 ? 不对, 应该直接是等于5/4乘以2, 即:
78
79
    b_4 + b_6 = 2*(5/4) = 5/2
80
81
82
    接下来,写出b₄和b。的表达式:
83
84
    b_4 = b_1 * q^{(4-1)} = b_1 * q^3
85
    b_6 = b_1 * q^{(6-1)} = b_1 * q^5
86
87
    所以,代入第二个条件:
88
89
    b_1*q^3 + b_1*q^5 = 5/2
90
91
92
    这里,我们可以用之前得到的方程(1)中的结果,即b_1*q^3 = 2。于是,代入到上面的式子:
93
```

 $2 + b_1*q^5 = 5/2$

```
这里需要处理一下。但根据方程(1),b_1*q^3 = 2,所以我们可以把b_1*q^5表示为b_1*q^3 * q^2 =
 96
     2*q2。因此,代入后:
 97
     2 + 2*q^2 = 5/2
98
99
     解这个方程:
100
101
     2*q^2 = 5/2 - 2 = 5/2 - 4/2 = 1/2
102
103
     所以q^2 = 1/4,因此q = 1/2,因为q是正数。
104
105
     接下来,代入方程(1)求b,:
106
107
108
     b_1*(1/2)^3 = 2 \rightarrow b_1*(1/8) = 2 \rightarrow b_1 = 2*8 = 16
109
     现在求前5项的和T。等比数列前n项和的公式是:
110
111
112
     T_n = b_1 * (1 - q^n)/(1 - q)
113
     代入n=5, b₁=16, q=1/2:
114
115
     T_5 = 16*(1 - (1/2)^5)/(1 - 1/2) = 16*(1 - 1/32)/(1/2) = 16*(31/32)/(1/2)
116
117
     计算分母1/2的倒数就是2,所以:
118
119
120
     16*(31/32)*2 = 16*(31/16) = 31
121
122
     所以T。的值是31,对应选项B。
123
     不过,检查一下步骤是否正确。特别是第二个条件那里,是否正确代入?
124
125
     原式是b4 + b6 = 5/2, 代入表达式:
126
127
     b<sub>1</sub>*q³ + b<sub>1</sub>*q^5 = 5/2。而根据方程(1),b<sub>1</sub>*q³ = 2,所以左边变为2 + 2*q²。代入后得到2
128
     + 2*q² = 5/2,解得q²=1/4,没错。之后计算正确。
129
     然后首项b,=16,计算前五项的和,得到31,对应选项B。看起来正确。"
130
131
             "答案": {
132
                "B"
133
134
                 }
135
        }
136
     }
```

6. 为什么选择数学 Why Math

数学作为逻辑推理和问题求解的核心学科,是衡量大语言模型理解能力与推理深度的重要标准。

更关键的是,尽管人工智能可以极大提升人类获取知识的效率,但它无法替代人类的思考。而逻辑能力是深度思考的基石,数学正是培养这一能力的最佳途径。因此,在教育类数据集中,数学数据集尤为重要,它不仅能提升 AI 的推理能力,还能让 AI 更好地辅助人类,帮助培养和强化逻辑思维能力。

7. 关于我们 About Us

上海临科智华数字科技有限公司是一家以技术为驱动的企业数据智能解决方案服务商。

我们的核心业务包括全流程的数据处理业务、大模型训推一体机和超算中心建设。

公司位于中国(上海)自由贸易试验区临港新片区 AI 产业园,注册资本金 3000 万元人民币。目前已与微软、百度、阿里、腾讯、京东、滴滴、字节跳动等国内外企业客户建立合作关系。

公司核心团队拥有十年以上人工智能数据服务经验,深刻理解多元业务场景的数据需求,可为自动驾驶、智能客服、智能制造、教育、智能医疗等多场景提供数据解决方案。

临科智华已在临港新片区建设全球领先的 AI 数据服务基地,以及具有业内引领性的人工智能辅助数据标注平台和一体化 AI 数据及资源管理平台,通过高质量的数据服务、专业的数据处理技术为中国(上海)自由贸易试验区临港新片区人工智能产业的快速发展提供坚实的基础支撑,加快推动 AI+多领域的创新与变革。

8. 使用许可 License

本数据集由上海临科智华数字科技有限公司(以下简称"授权方")发布,授权遵循CC BY-NC 4.0 (署名-非商业性使用)协议,并作如下补充说明:

8.1 非商业用途授权

- 本数据集仅授权用于学术研究、教学及非商业目的的使用。
- 用户可在署名原作者的前提下自由使用、共享、改编本数据集,但不得用于商业用途。
- 使用本数据集进行研究或发表成果时,必须在显著位置注明数据来源及引用本数据集的原始发布者。

8.2 商业用途限制

未经授权方书面许可,任何单位或个人不得将本数据集用于任何商业目的,包括但不限于:

- 将数据用于付费产品或服务;
- 用于公司内部模型训练、产品开发;
- 以数据为基础开展商业合作等。

如需将本数据集用于商业用途,请联系授权方申请书面授权。授权方保留根据使用场景收取授权费用的权利。

8.3 免责声明

- 授权方不对数据内容的准确性、完整性或适用性承担任何明示或暗示的责任。
- 用户在使用过程中应遵守所在国家和地区的法律法规,由用户自行承担使用过程中的一切法律风险。

9. 联系我们 Contact Us

上海临科智华数字科技有限公司

地址:上海市浦东新区临港新片区港城广场一街坊9号楼

电话: 021-60701336