

数学应用题生成流程 V1 & V1.5

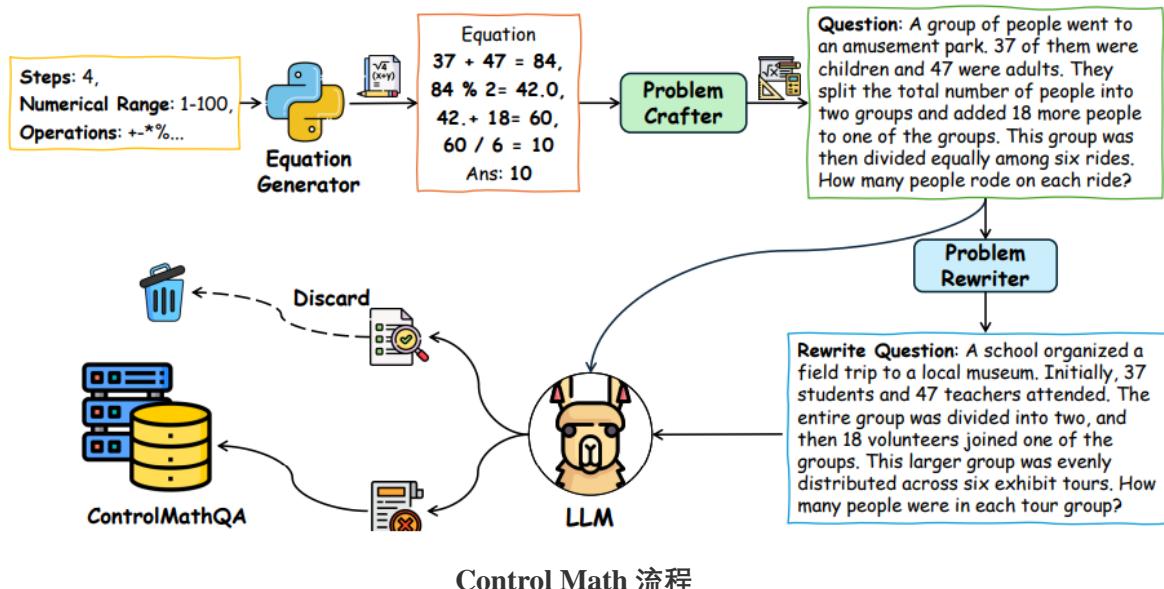
1. 引言

- **背景**：physics 2.1 和 controlMath 中提到基本数学运算符出发，合成长运算逻辑，长变量依赖关系的数学问题
- **目的**：提高现有模型数学能力及推理能力需要大量高质量数学&推理数据，而当下大部分高质量数学数据都是来源于 GSM8K 的变体。所以我们尝试探索是否存在一种可以用更低的成本去批量生成高质量数学数据的流程。
- 参考文章：
 - Physics 2.1:<https://arxiv.org/pdf/2407.20311>
 - ControlMath: [\[2409.15376\] ControlMath: Controllable Data Generation Promotes Math Generalist Models](#)

2. 数学应用题生成流程 1.0：

2.1 初期思路：

2.1.1 参考 ControlMath



- ControlMath 框架在构造上十分简洁便利，它要求模型基于给定的解答过程发散生成题目。

而需要准备的数据仅仅是随机几行数学式。

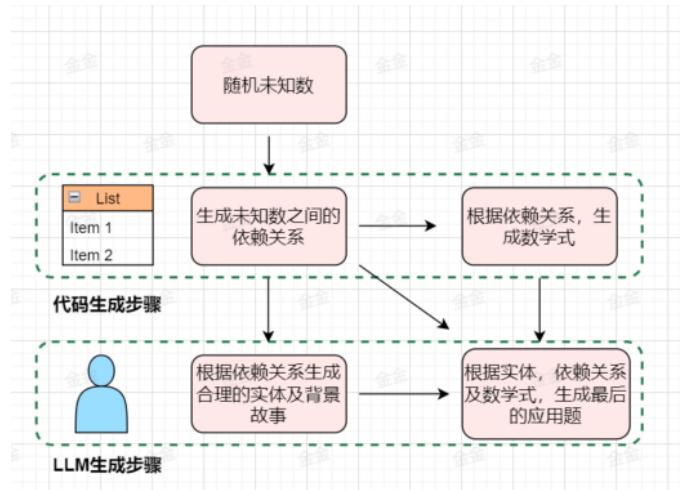
- 但经过复现，直接基于数学式生成的题目有以下几个问题：
 - 题目情境重复率极高。70%左右的题目集中在志愿者、书店、节日嘉年华这三大故事中
 - 生成的解答过程不连贯（前面的式子算出来的答案应当会成为后续计算过程的条件，不然就是没有意义的条件）
 - 例如，我们随机出的第一条数学式为 $3+9=12$ ，而“12”这个计算答案在后续的解题过程中再也没出现过。所以对于这个解题过程，“ $3+9=12$ ”这个式子就是多余的。
 - 题目给到的条件单一，且部分不符合常理（例如总人数小于成年人人数，或出现 $1/2$ 个人）

2.1.2 参考 Physics 2.1

- Physics 2.1 的框架在构造上极其繁琐，且存在短时间难以处理的问题（例如同一数值的连续性），但它的思路可以借鉴并用来改进 ControlMath 的流程。
- 针对题目情境重复的问题，我们改进了模型生成的流程，包括让模型：
 - 生成预设主题；
 - 生成针对该主题的背景故事；
 - 生成该背景故事下可能出现的、针对每一个实体的现实含义。
- 在得到以上信息后，再让模型整合出最后的应用题。
- 实体的含义：实体在应用题里的内容，可以是条件的一部分，也可以是被提问的未知数，例如某个房间的面积，某个家具的价钱

针对以上两个流程，我们提出了以下增强版流程：

2.2 题目生成流程：



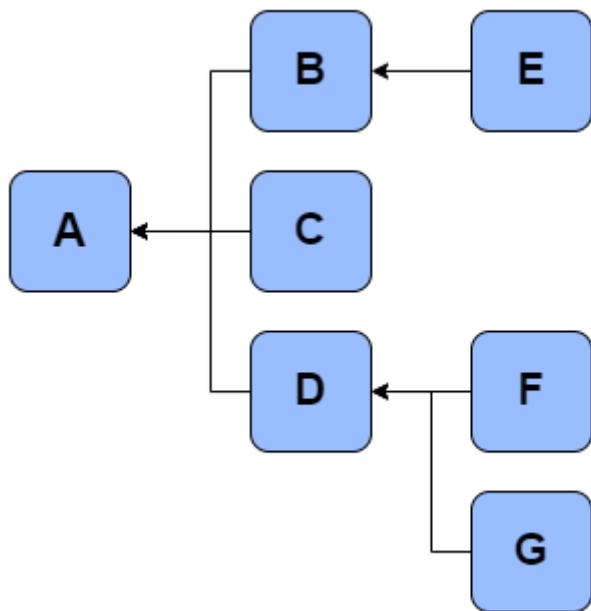
2.2.1 第一步：随机未知数

格式为大写英文字母，如 A, B, C。仅作为在设定其含义前出现的占位符使用。

2.2.2 第二步：生成未知数之间的依赖关系

以下为一个可能的依赖关系案例：

如 A 需要 B, C, D 来计算，B 需要 E 来计算，D 需要 F, G 来计算：



依赖结构必须为一个树结构，内部不存在循环，这样我们根据依赖图生成的计算过程就是最优解

2.2.3 第三步：根据依赖关系随机生成计算过程

根据上一个依赖关系案例，我们可能随机出以下关系：

$$A = B + C + D \quad A \text{ 是 } B, C, D \text{ 的总和}$$

$$B = E * 2 \quad B \text{ 是 } E \text{ 的两倍}$$

$$D = F - G \quad D \text{ 是 } F \text{ 减去 } G$$

我们用实体之间的依赖关系替代随机数学式作为地基，生成符合逻辑的实体和数学运算片段，再生成对应的答案。用这个方式生成的计算过程可以做到每个树权上的实体都会作为条件参与进计算，不会凭空出现没用的计算式。这样就解决了“生成的解答过程不连贯”这个问题。

2.2.4 第四步：生成主题灵感

基于常见应用题类型，我们收集了以下主题：

- 行为类，例如：分数计算、空间分配
- 单位类，例如：面积、距离、收益
- 名词类，例如：玩具、零食、主题乐园
- ...

一共 500 个内容完全不重叠的主题，作为题目的灵感启发片段

2.2.5 第五步：生成背景故事&实体含义

根据依赖关系使用模型生成依赖关系匹配逻辑的背景故事和实体含义

Prompt:

You are an AI model specializing in creating captivating and educational primary school math word problems. Your task is to generate a scene for a math application problem. That is, you need to define the literal meaning of each given variable using the provided their dependencies, known values, and the solution expressed as equations. Your scene set will be used to generate a problem that not only challenges young learners but also relates to real-world scenarios.

Below is a sample input and corresponding output for your reference:

Sample Input:

```
topic:{{greening area of city}}
variable_symbols: {{A, B, C, D, E, H}}
variable_dependency: {{A dependents on B and C, B dependents on D and E, E dependents on H}}
variable_mathematical_relationship: {{A is the sum of B and C, B is the sum of D and E, E is the half of H}}
known_value: {{C, D, H}}
variable you want to ask about: A
```

Sample Output:

[Generation]:

```
{ {
    "Variable": {{A = 'Exceeded greening area of the city' , B = 'piratical greening area of the city' , C = 'Expected greening area of the city', D = 'greening area of District B', E = 'greening area of District A', H = 'Area of District A'}}
    "Background_story": 'In order to cope with the inspection of the superiors, our city needs to do more statistics on the green area'
} }
```

The final Question generation for you to look at:

In the city planning office, a new initiative was launched to increase the greening area across various districts of the city. This initiative involved measuring areas and ensuring that each district met its greening goals.

District A, known for its sprawling parks, had a total area of 44 square kilometers. The city planners decided to allocate half of this area specifically to green spaces. Meanwhile, District B was noted to have already established 13 square kilometers dedicated to greenery.

The city's goal was to reach a total of 32 square kilometers of greening area across all districts. Considering these plans and

该 prompt 必要片段为：

1. few shot 案例：是我们基于一个依赖关系想出来的理想案例，该题目是：
 - a. 在城市规划办公室，一个新的计划被启动，目的是增加城市各个区域的绿化面积。该计划包括测量各区域的面积，并确保每个区域都能达到其绿化目标。A 区以其大片公园而闻名，总面积为 44 平方公里。城市规划者决定将这一面积的一半专门用作绿地。与此同时，B 区已明确有 13 平方公里用于绿化。城市的目标是使所有区域的绿化总面积达到 32 平方公里。考虑到 A 区和 B 区已分配的绿化面积以及新的规划调整，城市在实施新计划后，其绿化面积目标是超额还是不足，以及具体的超额或不足数值是多少？
 - b. 此案例中的每个实体的含义与其互相的依赖关系匹配程度很高，且非常合理，通顺。是日常生活中会遇到的数学应用题。
 2. 步骤规划：
 - a. 基于给定主题联想合理的背景故事，确保实体之间的数学关系（即：加、减、乘、除）能被合理地囊括在该背景故事之中
 - b. 定义每个实体的含义
 - c. 基于实体依赖关系对实体设定好的含义作检查，若出现不合理的条件，重新修改该实体的含义
- 例子：
 - sample input:

```
{  
    "variable_symbols": "A, B, C, D, E, F, G, H, I, J",  
    "variable_dependency": "A dependents on B, C, D, E and F, B  
dependents on G and H, C dependents on I, D dependents on J",  
    "variable_combined_relationship": "A is the sum of B, C, D, E  
and F, B is the multiplication of G and H",  
    "variable_mathematical_relationship": "D is J divided by four,  
C is five times I",  
    "known_value": "E, F, G, H, I, J",  
    "known_value_with_data": "E = 64, F = 7, G = 7, H = 40, J =  
12, I = 5",  
}
```

必要输入为：题目主题、实体、实体之间的依赖关系、实体是否已知。

输出如下所示：

- sample output:

You are an AI model specializing in creating captivating and educational primary school math word problems. Your task is to generate a single math application problem using the provided variable symbols, their dependencies, known values, and the solution expressed as equations. The end goal is to craft a problem that not only challenges young learners but also relates to real-world scenarios.

Below is a sample input and corresponding output for your reference:

Sample Input:

background story: 'In an ambitious urban sustainability initiative, a city's authorities set a target to enhance its green spaces'

variable symbols: {{A = 'Exceeded greening area of the city' , B = 'piratical greening area of the city' , C = 'Expected greening area of the city', D = 'greening area of District B', E = 'greening area of District A', H = 'Area of District A'}}}

variable dependency: {{E dependents on H, B dependents on D and E, A dependents on B and C, B is sum of E and D, A is B minus C, H half is E}}

known value with data: C = 32, D = 13, H = 44

variable you want to ask about: A

Sample Output:

In the city planning office, a new initiative was launched to increase the greening area across various districts of the city. This initiative involved measuring areas and ensuring that each district met its greening goals.

Here you know:

1. District A, known for its sprawling parks, had a total area of 44 square kilometers. The city planners decided to allocate half of this area specifically to green spaces.
2. piratical greening area of the city is composed of greening area of District A and greening area of District B
3. Meanwhile, District B was noted to have already established 13 square kilometers dedicated to greenery.
4. The city's goal was to reach a total of 32 square kilometers of greening area across all districts
5. The difference between 'piratical greening area of the city' and 'Expected greening area of the city' is Exceeded greening area of the city

Finally I want to know what is Exceeded greening area of the city?

这个 prompt 的目的是让 模型根据给定的背景故事、变量符号、变量之间的依赖关系、已知值以及要计算的变量，生成一个具有吸引力和教育意义的数学应用题。

重点在于：

1. 利用给定的信息，将数学依赖关系转化为现实世界中的叙事元素，创建一个连贯的问题。
2. 确保在问题中清晰地表达数学依赖关系，不使用字母符号，而是用明确的文字描述。
3. 整合所有已知条件，形成一个完整的问题陈述，使所有已知条件都出现在问题中。
4. 使用明确的方程和数字来说明每个变量是如何计算的，使问题既有趣又具有教育意义。

2.2.6.1 题目生成示例

输入：

```
{  
    "background_story": "{'In a bustling town, the local  
community center is organizing a grand festival. The festival  
will feature various activities and attractions, and the  
organizers need to calculate the total number of participants  
expected. They have different groups contributing to the total  
number of participants, including school children, local clubs,  
and neighborhood associations. Each group has its own way of  
estimating the number of participants.'}\n        }",  
    "variable_entity": "{A = 'Total number of participants  
expected at the festival', B = 'Number of participants from the  
school children group', C = 'Number of participants from the  
local clubs', D = 'Number of participants from the neighborhood  
associations', E = 'Number of participants from the sports  
teams', F = 'Number of participants from the arts and crafts  
groups', G = 'Number of schools participating', H = 'Average  
number of children per school', I = 'Number of local clubs  
participating', J = 'Number of neighborhood associations  
participating'}," ,  
    "variable_dependency": "A dependents on B, C, D, E and F,  
B dependents on G and H, C dependents on I, D dependents on J",  
    "variable_combined_relationship": "A is the sum of B, C,  
D, E and F, B is the multiplication of G and H",  
    "variable_mathematical_relationship": "D is J divided by  
four, C is five times I",  
    "known_value_with_data": "E = 64, F = 7, G = 7, H = 40, J  
= 12, I = 5"}  
}
```

输出:

question : In a bustling town, the local community center is organizing a grand festival. The festival will feature various activities and attractions, and the organizers need to calculate the total number of participants expected. They have different groups contributing to the total number of participants, including school children, local clubs, neighborhood associations, sports teams, and arts and crafts groups. Each group has its own way of estimating the number of participants.

Here is what we know:

1. There are 7 schools participating in the festival, and each school is expected to bring an average of 40 children.
2. There are 5 local clubs participating, and each club is expected to bring 5 times the number of participants.
3. There are 12 neighborhood associations participating, and each association is expected to bring a quarter of its members.
4. The sports teams are expected to bring 64 participants.
5. The arts and crafts groups are expected to bring 7 participants.

Considering these estimates, can you calculate the total number of participants expected at the festival?

solution : To find the total number of participants expected at the festival, we need to calculate the number of participants from each group and then sum them up.

1. Number of participants from the school children group (B) :

$$\begin{aligned} B &= G * H \\ B &= 7 * 40 \\ B &= 280 \end{aligned}$$

2. Number of participants from the local clubs (C) :

$$\begin{aligned} C &= 5 * I \\ C &= 5 * 5 \\ C &= 25 \end{aligned}$$

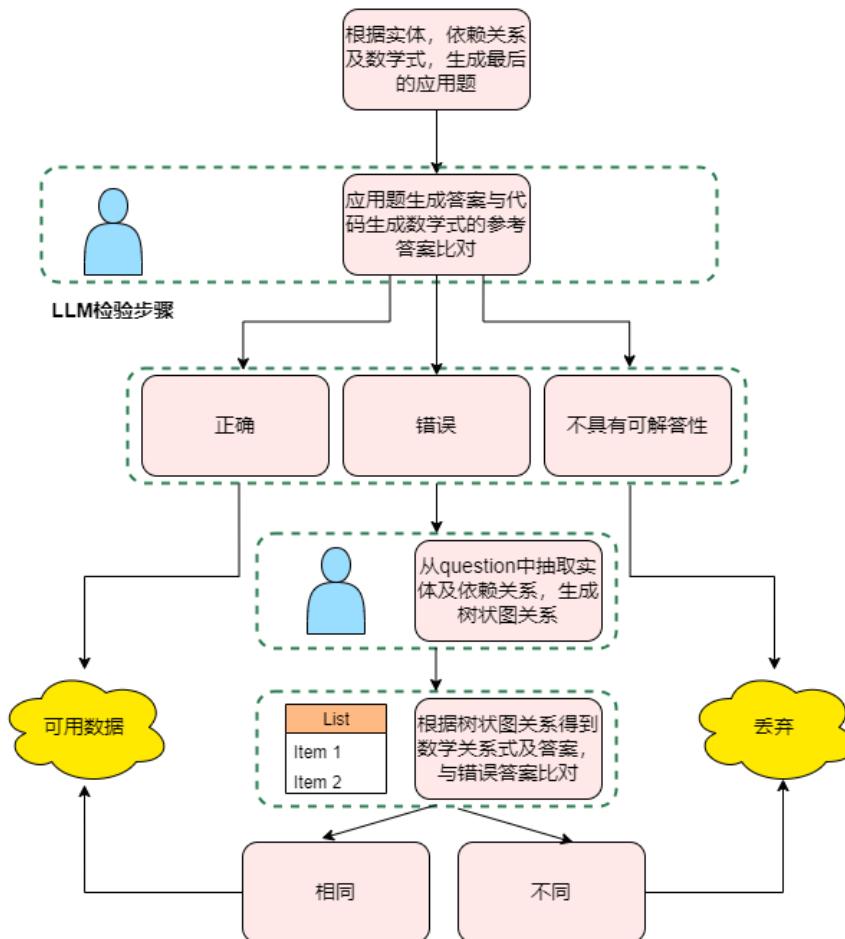
3. Number of participants from the neighborhood associations (D) :

$$\begin{aligned} D &= J / 4 \\ D &= 12 / 4 \\ D &= 3 \end{aligned}$$

4. Number of participants from the sports teams (E) :

在生成完整问题后，我们发现问题的真实答案仍然会与我们给到的输入答案有一定差异，所以一个根据现有问题去生成答案的流程是必要的

2.4 答案生成流程：



答案生成流程如图所示

2.4.1 第一步：题目分类

首先，使用大模型判断数据是否可解。对于可解数据，做出解答后与原答案进行比对。若答案相同，则该答案视作正确答案。若答案不同，则暂时没有确定答案。至此，我们的数据分为三个分类：题目不可解、题目可解且有正确答案、题目可解但没有正确答案。

对于不可解的题目数据（约占总数据的 5%），我们选择直接丢弃。对于可解且有正确答案的

(约占总数据的 60%)，我们可以直接录用。接下来我们只需要处理题目可解但没有正确答案的分类。

2.4.2 第二步：重做错误题目获取正确答案

对于题目可解但没有正确答案的数据，我们增加了以下流程来获取正确答案：

将题目交由大模型分析其逻辑框架，再转换成程序语言，计算其最终答案。

You are an insightful problem-solver equipped with an exceptional ability to identify and articulate variable dependencies within complex logical scenarios. Your goal is to accurately interpret the relationships between variables from a text-based description and to present them in a structured, hierarchical format that clearly maps out these dependencies.

Your task is to analyze a given problem description and determine the tree of dependencies for the variables involved. This involves identifying variables, their values, and the relationships between them as described in the problem. You should present your findings in a clear, structured format that highlights the dependencies.

Sample Input:

In a school drawing competition, students are excited to showcase their artistic talents. Each student submits a certain number of drawings, and the competition organizers keep track of the total number of drawings submitted. Jack, Bella, David, and Sarah are some of the participants, and their contributions are recorded to determine the overall success of the competition.

Here is what we know:

1. Jack submitted 27 drawings.
2. Bella submitted 14 drawings.
3. Sarah submitted 7 drawings.
4. There are a total of 22 participants in the competition.
5. The average number of drawings submitted per participant is calculated by dividing the total number of participants by 11.
6. The total number of drawings submitted by all participants except Jack is 33 times the average number of drawings submitted per participant.
7. The total number of drawings submitted by participants excluding Jack and Bella is the sum of the total number of drawings submitted by all participants except Jack and the number of drawings submitted by Jack.
8. David submitted 25 more drawings than the total number of drawings submitted by participants excluding Jack and Bella.
9. The total number of drawings submitted by Jack, Bella, and David is the sum of the number of drawings submitted by David and Bella.
10. The total number of drawings submitted in the competition is

经过处理后单条错误数据的信息汇总样例，包括模型输出：

"topic": "Toy",
"question": "Customer footfall in Ninja Turtles per day is 8.
Customer footfall in Minecraft per day is 12.
Customer footfall in Minions per day is 10.
Customer footfall in DC Comics per day is 1.
Customer footfall in Avatar: The Last Airbender per day is 5.
Customer footfall in Toy Story per day is 4.
Customer footfall in Barbie per day is 3.
Customer footfall in Incredibles per day is 4.
Customer footfall in Trolls per day is sum of customer footfall
in Jurassic Park per day and customer footfall in Minecraft per
day .
Customer footfall in Spider-Man per day is difference of customer
footfall in The Incredibles per day and customer footfall in
Barbie per day .
Customer footfall in Moana per day is customer footfall in My
Little Pony per day divided by 2 number .
Customer footfall in My Little Pony per day is difference of
customer footfall in Star Wars per day and customer footfall in
Minions per day .
Customer footfall in Paw Patrol per day is 7 more then customer
footfall in Dora the Explorer per day .
Customer footfall in The Incredibles per day is 12 times as much
as customer footfall in Toy Story per day .
Customer footfall in Batman per day is 9 more then customer
footfall in Moana per day .
Customer footfall in Dora the Explorer per day is difference of
customer footfall in Ninja Turtles per day and customer footfall
in Avatar: The Last Airbender per day .
Customer footfall in Star Wars per day is 3 more then customer
footfall in Stranger Things per day .
Customer footfall in Jurassic Park per day is sum of customer
footfall in Batman per day and customer footfall in Incredibles
per day .
Customer footfall in Monsters Inc. per day is 4 less then
customer footfall in Paw Patrol per day .
Customer footfall in Stranger Things per day is difference of
customer footfall in Marvel per day and customer footfall in DC
Comics per day .
Customer footfall in Rick and Morty per day is 43 less then
customer footfall in Spider-Man per day .
Customer footfall in Marvel per day is 24 more then customer
footfall in Looney Tunes per day .

再经过“`Relationships_in_Equation_Forms`”和“`Hierarchical_Dependency_Mapping`”对最终答案进行计算。若该答案与模型生成答案一致，则判定为正确答案。若否，则丢弃该数据。

2.5 完整流程结果示例：

2.5.1 生成数据效果：

生成数据格式参考

"topic": "Graduation Day",
"question": "It's Graduation Day at Sunshine Elementary School, and the school administration is busy preparing for the big event. They need to ensure that they have enough graduation certificates for all the students who are eligible to graduate. The principal, Ms. Green, is working with her team to calculate the total number of certificates needed. They know that the number of students eligible for graduation depends on several factors, including the number of students who completed all courses, those who received honors, and those who transferred or dropped out. Ms. Green has gathered data on the total number of students in the school, the number of students who enrolled in the graduation ceremony, and the number of students who failed to meet graduation requirements. With this information, they can determine how many certificates to prepare for the ceremony. Here is the information they have gathered:

1. The total number of students in the school is 20.
2. The number of students who enrolled in the graduation ceremony is 26.
3. The number of students who transferred to another school is 5.
4. The number of students who failed to meet graduation requirements is 1.
5. The number of students who completed all courses is 40 more than the number of students who dropped out.
6. The number of students who received honors is half of the total number of students in the school.
7. The number of students who dropped out is the difference between the number of students who enrolled in the graduation ceremony and the number of students who failed to meet graduation requirements.
8. The number of students eligible for graduation is the difference between the total number of students who completed all requirements and the number of students who transferred to another school.
9. The total number of students who completed all requirements is the sum of the number of students who completed all courses and the number of students who received honors.
10. The total number of graduation certificates needed is 50 more than the number of students eligible for graduation.

Based on this information, can you determine the total number of graduation certificates needed?",

"answer_list": [" $120 = 70 + 50$ ", " $65 = 25 + 40$ ", " $10 = 20 / 2$ ", " $25 = 26 - 1$ ", " $70 = 75 - 5$ ", " $75 = 65 + 10$ "],

2.5.2 数据统计：

数据名	树深	树宽	正确答案/错误答案/无解/总数据量
question_2w_10_3_7.json	7	3	3002/1550/402/5000
question_2w_10_3_9.json	≤ 9	≤ 3	1355/409/135/2000
question_2w_10_9_9.json	≤ 9	≤ 9	3650/1044/312/5000
question_2w_10_5_5.json	≤ 5	≤ 5	1341/538/103/2000
question_2w_11_3_9.json	≤ 9	≤ 3	1533/865/1034/3500
question_1w_10_3_9.json	≤ 9	≤ 3	6213/2259/1162/10000
question_1w_11_3_9.json	≤ 9	≤ 3	3058/1167/756/5000
question_2w5_10_3_9.json	≤ 9	≤ 3	3406/1345/843/5605
question_2w5_11_3_9.json	≤ 9	≤ 3	30294/11774/6742/50000
question_2w_12_3_9.json	≤ 9	≤ 3	11871/4733/3319/20000

2.6 小结

- 总结：

1. 数据可以达到总生成量的 70%左右的利用率
2. 从 error 中提取的 new_right 的成功率为 40%左右，对于利用率的提升有较大帮助
3. 经验证，从 error 中提取的 new_right 对应 question 在 4O 上解答正确率不高，可见重筛选的数据所覆盖的逻辑甚至超过 4O 本身逻辑推理能力

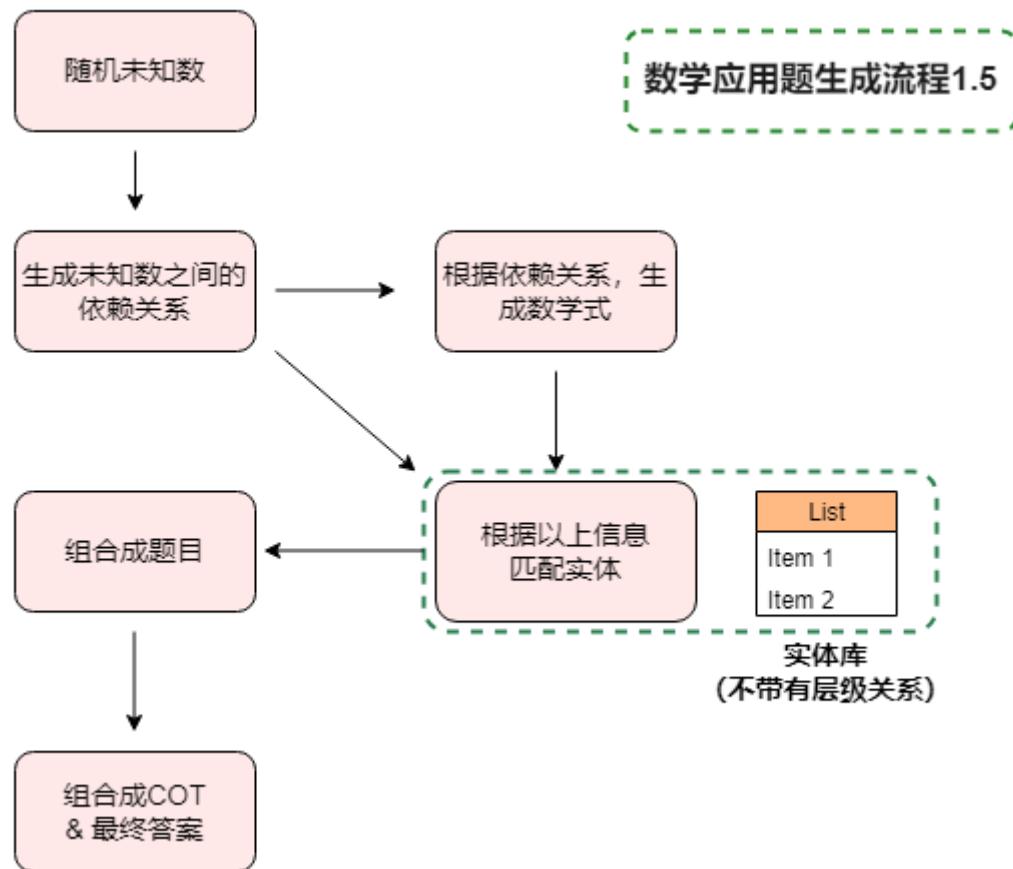
待优化项：

- 由于实体含义是靠模型生成的，它存在难度上限。当实体数量大于 10 的时候，模型输出的实体含义就会出现不合理的设定：
 - 例如某个未知数 $A=B-C$ ，模型将 B 的含义设定为科研组人员数量， C 的含义设定为科研部门人员数量， A 的含义设定为技术组人员数量。
- 模型的解答能力也会随着未知数数量的上涨而下降，尤其是依赖关系逐渐复杂的情况下。
- 依赖关系越复杂，参与差值条件和除法条件的实体越难被合理定义。
 - 例如存在 $A=B-C$ ，我们可以将 A 定义为 B 和 C 的差值。但如果 A 后续又参与了另一个除法或减法，那这个条件里的实体就很难被合理定义了
 - 而简单地调控每个实体的被定义顺序无法解决这个问题，因为在定义单个实体时，我们需要同时知道它参与的两个条件是否存在差值与除法。理论上被减数和被除数的定义域应当要比差和商的定义域宽广，差应当是被减数的子集，商应当是被除数的单位/单位乘数（后续 3.0 会提到的，前者的层数应当比后者高）

3. 数学应用题生成流程 V1.5:

3.1 初期思路：

为了解决当前生成的题目难度有上限的问题，我们尝试将模型生成合理实体含义的步骤改成预生成实体库，并为每个实体匹配同一场景下的含义。



```

"Snacks": [
    67,
    {
        "choices": 63,
        "entity" : [
            "Potato chips", "Popcorn",
            "Pretzels", ... ],
            ["Snickers", "Twix", "M&M's", ...],
            ["water", "coffee", "tea", "milk", ...],
            ["salty flavored", "sweet flavored", ...],
            ["single-serve", "family-size",
            "resealable", ...],
            ["whole grain", "gluten-free",
            "organic", ...],
            ["everyday", "on-the-go", "special
            events", ...],
            "key": ["Cost differential for {} product",
            "Market demand for {} product", "prize of a pack of {} product",
            f"number of {{}} product sold per {random.choice(frequency)}",
            "average calorie count of {} product", "popularity level of {}"
            product", "consumer feedback rating on {} product", "sales
            increase due to new {} product", "number of new released {}"
            product in market"]
    }
]

```

如上所示，实体条目应包含：

- 题目主题：“`Snacks`”
- 实体列表：“`entity`”
- 针对以上实体的可提问方式：“`key`”

如示例中，针对列表：爆米花、薯片、山药片等零食，我们可以提问：市场需求量，单价，热量等。

在生成依赖关系后，为当前题目在指定主题中随机一个实体列表和提问方式，并为实体在列表中随机含义，最后再拼凑成完整题目

如此生成的题目突破了实体数量的上限，但语句不是很通顺。

1.5 生成的问题如下：

"鳕鱼区的入场费是 3。\\n 海豹区的入场费是 1。\\n 黑线鳕区的入场费是 10。\\n 鹦嘴鱼区的入场费是 1。\\n 鼠海豚区的入场费是 14。\\n 水螅区的入场费是 2。\\n 墨鱼区的入场费是 13。\\n 螃蟹区的入场费是 4。\\n 双髻鲨区的入场费是狗鱼区的入场费减去海豹区的入场费。\\n 鲨鱼区 60 的入场费比海星区的入场费多 21。\\n 飞鱼区 26 的入场费比硬壳蛤区的入场费多 22。\\n 狗鱼区的入场费是海獭区的入场费加上螃蟹区的入场费。\\n 蓝鲸区 -14 的入场费是鲱鱼区的入场费减去鼠海豚区的入场费。\\n 虎鲸区 6 的入场费是鲨鱼区的入场费除以 10。\\n 海星区 39 的入场费是墨鱼区的入场费的 3 倍。\\n 鲱鱼区 3 的入场费是海绵区的入场费减去鹦嘴鱼区的入场费。\\n 海绵区 4 的入场费是海豹和鳕鱼区的入场费之和。\\n 海獭区的入场费是蓝鲸区的入场费除以 3。\\n 海象区 5 的入场费是黑线鳕区的入场费除以 2。\\n 硬壳蛤区 4 的入场费是虎鲸区的入场费减去水螅区的入场费。\\n 海豹区 21 的入场费是飞鱼区的入场费减去海象区的入场费。\\n\\n 计算双髻鲨区的入场费。"

1.0 的问题如下：

"在一所小学，学生们正在进行一个涉及多个团队的大型项目：艺术、科学、数学、历史、体育和音乐。每个团队负责完成特定任务，以推动项目整体进展。校长希望跟踪每个团队的进展以及项目的总体进展。艺术团队、科学团队和数学团队做出了显著贡献，而历史团队和体育团队也在努力工作。校长为要完成的任务总数设定了一个目标，并想了解目前的进展情况。以下是我们所知道的：

科学团队已完成 10 个实验。体育团队的进展相当于 6 个任务。数学团队已完成 12 个任务。音乐团队的进展相当于 14 个任务。历史团队已完成 3 个任务。校长设定的初始任务数量是 10 个。数学团队和历史团队的总进展是 9 个任务。

校长还知道以下几点：

完成的任务总数比初始任务数多 19 个。剩余要完成的任务数比已完成任务总数少 21 个。剩余任务的一半是总进展计算的一部分。艺术团队的进展是科学、数学和历史团队综合进展以及体育团队进展之和。数学团队和历史团队之间的进展差异是数学和历史团队综合进展与历史团队进展之间的差异。学校项目的总进展是剩余任务的一半、艺术团队的进展和音乐团队的进展之和。

你能计算出学校项目的总进展吗？"

3.3 效果反馈：

- 生成问题的利用率可以达到 75%
- 训练的实验效果如下：

模型名称	训练集	测试集	准确率 = correct / total
tianji-2b-v9-dpo	未经训练	op10	$3659 / 10309 = 0.3549$
tianji-2b-v9-dpo	未经训练	op20	$1 / 1665 = 0.0$
tianji-2b-v9-dpo	未经训练	op10+op20	$3667 / 11974 = 0.3062$
tianji-2b-v9-base	op10+op20 (107394)	op10	$8317 / 10309 = 0.8068$
tianji-2b-v9-base	op10+op20 (107394)	op20	$1160 / 1665 = 0.6967$
tianji-2b-v9-base	op10+op20 (107394)	op10+op20	$9482 / 11974 = 0.7919$
qwen2.5-7b-math	-	-	0.70

4.2 改进方向：

- 所有存在运算关系的实体在自然语言逻辑上自洽且符合常理

“小明的运动时间是 5，小明的运动速度是 3，小明的运动距离等于小明的运动时间乘以小明的运动速度，小红的运动距离是小明运动距离的两倍”

而非硬凑逻辑关系：

“数学团队和历史团队之间的进展差异是数学和历史团队综合进展与历史团队进展之间的差。学校项目的总进展是剩余任务的一半、艺术团队的进展和音乐团队的进展之和”

- 合成的题目能够省略掉常识计算法则

可以删去：“小明的运动距离等于小明的运动时间乘以小明的运动速度”

4.2.2 改进流程：

- 进一步升级实体库，使实体库中的实体内嵌复杂的上下级结构关系
- 改进实体树搭建流程。令生成的题目条件包含实体含义之间的条件关系，而不只是数值上的关系