SciDA: Scientific Dynamic Assessor of LLMs

Data Collection Workflow

Data Collection

Data Source
- Olympiads
- University Textbooks
- Private & Original Problems

Olympiads Award-Winning Students

Abandoned Data

Data Filtering

Problem Segmentation Team

Selected Data

For Further Processing

Filter Criteria
- Sufficient Difficulty
- Numerical Calculation Problem
- Answer Determined by Variables

Data Annotation

Specialized Annotation Team

Annotation Process

x **Annotate variables**: Enclosing them within "\\$" symbols

$f(x)$ **Functionalize problems**: Covert them into python code

---

Disciplines

1. Mathmatics
   a. Calculus   b. Algebra   c. Number Theory

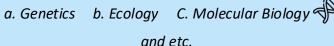**2. Physics**
   a. Mechanics   b. Optics   c. Astrophysics
   d. Electrodynamics   e. Quantum Mechanics
   f. Thermodynamics and Statistical Physics

3. Chemistry
   a. Physical Chemistry   b. Analytical Chemistry
   c. Inorganic chemistry   d. Organic Chemistry

4. Biology
   a. Genetics   b. Ecology   C. Molecular Biology
   and etc.

Problem Paradigm

1. Variable parameters

Variables: $\{ X_1, X_2 ..., X_J \}$

X_1 range: (a_1, b_1)

...

X_j range: (a_j, b_j)

The parameters are allowed to vary In certain range

2. Numerical Answer

Answer: $Y = F(X_1, X_2, ..., X_J)$

The answer can be
   a. functionalized
   b. calculated in numerical form

Dynamic Random Initialization

Source

1. Olympiads
2. University Textbooks
3. Private & Original Problems

Vision

Mitigate data contamination Toward truthful assessments