JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

https://promptpg.github.io

Dynamic Prompt Learning via Policy Gradient for Semi-structured Mathematical Reasoning

Pan Lu[1,3], Liang Qiu1, Kai-Wei Chang[1], Ying Nian Wu[1], Song-Chun Zhu[1], Tanmay Rajpurohit[2], Peter Clark[3], Ashwin Kalyan[3]

[1]University of California, Los Angeles [2]Georgia Institute of Technology [3]Allen Institute for AI

UCLA Georgia Tech AI2

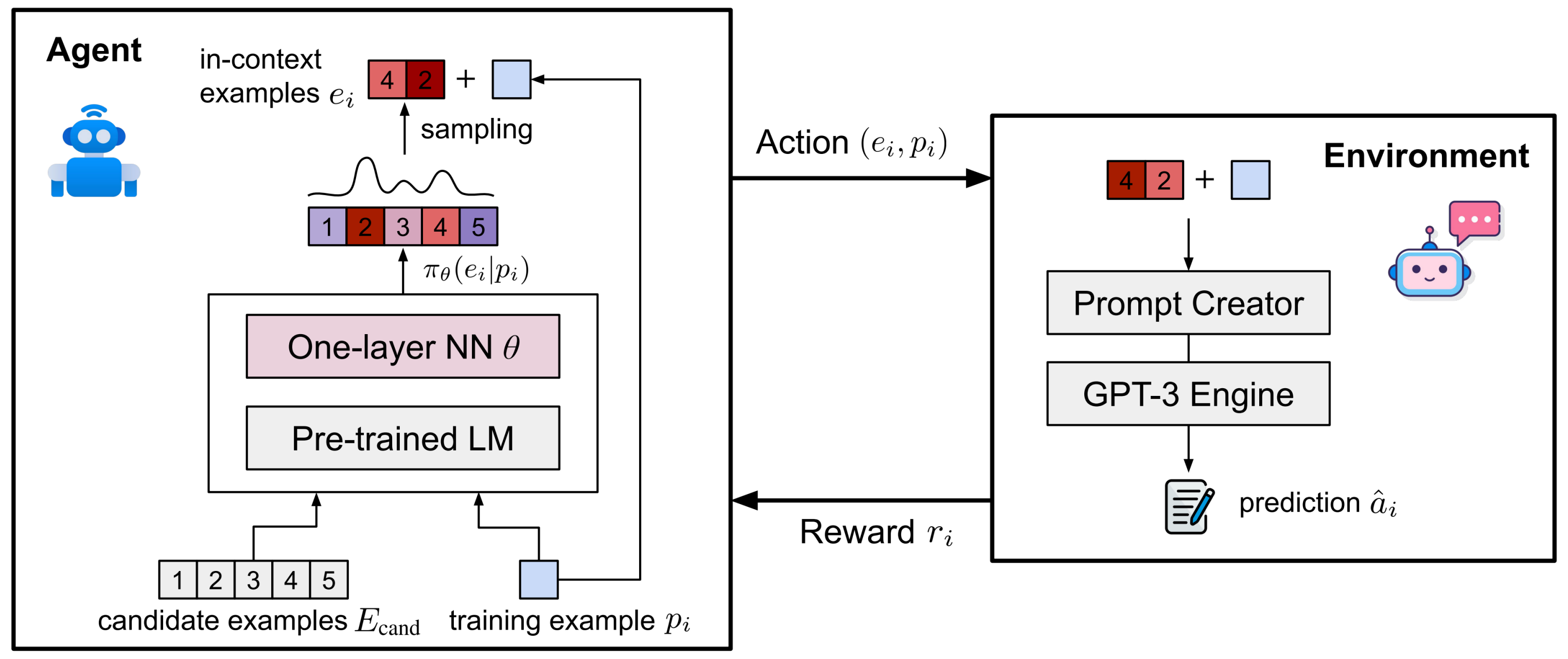
Introduction

square beads	\$2.97 per kilogram
oval beads	\$3.41 per kilogram
flower-shaped beads	\$2.18 per kilogram
star-shaped beads	\$1.95 per kilogram
heart-shaped beads	\$1.52 per kilogram
spherical beads	\$3.42 per kilogram
rectangular beads	\$1.97 per kilogram

Question: If Tracy buys 5 kilograms of spherical beads, 4 kilograms of star-shaped beads, and 3 kilograms of flower-shaped beads, how much will she spend? (unit: \$)
Answer: 31.44
Solution:
Find the cost of the spherical beads. Multiply: \$3.42 × 5 = \$17.10.
Find the cost of the star-shaped beads. Multiply: \$1.95 × 4 = \$7.80.
Find the cost of the flower-shaped beads. Multiply: \$2.18 × 3 = \$6.54.
Now find the total cost by adding: \$17.10 + \$7.80 + \$6.54 = \$31.44.
She will spend \$31.44.

Sandwich sales		
Shop	**Tuna**	**Egg salad**
City Cafe	6	5
Sandwich City	3	12
Express Sandwiches	7	17
Sam's Sandwich Shop	1	6
Kelly's Subs	3	4

Question: As part of a project for health class, Cara surveyed local delis about the kinds of sandwiches sold. Which shop sold fewer sandwiches, Sandwich City or Express Sandwiches?
Options: (A) Sandwich City (B) Express Sandwiches
Answer: (A) Sandwich City
Solution:
Add the numbers in the Sandwich City row. Then, add the numbers in the Express Sandwiches row.
Sandwich City: 3 + 12 = 15. Express Sandwiches: 7 + 17 = 24.
15 is less than 24. Sandwich City sold fewer sandwiches.



- We propose **TabMWP**, the first dataset for **math word problems with tabular context**
- We propose **PromptPG**, the first work that applies **reinforcement learning** to select in-context examples for the few-shot GPT-3 model

Tabular Math Word Problem (TabMWP) Dataset

2 Tasks **38,431** Problems **35,442** Solutions **37,644** Tables **12.9/54** Avg/Max cells

Statistic	Number
Total questions	38,431
* *free-text* questions	28,719
* *multi-choice* questions	9,712
# of different questions	28,876
# of different answers	6,153
# of different solutions	35,442
# of different tables	37,644
# of tables with a title	23,259
# of table cells (Average/Max)	12.9 / 54
# of table rows (Average/Max)	5.9 / 11
# of table columns (Average/Max)	2.2 / 6
Question length (Average/Max)	22.1 / 92
Answer length (Average/Max)	1.1 / 27
Solution length (Average/Max)	49.5 / 350

- It contains **38,431** open-domain grade-level problems that require **mathematical reasoning** on both textual and tabular data
- Each question in TabMWP is aligned with a **tabular context**, which is presented as an image, semi-structured text, and a structured table
- There are two types of questions: **free-text** and **multi-choice**
- Each problem is annotated with gold solutions to reveal the **multi-step** reasoning process

Dataset	Size	#Table	Need Math?	Need Table?	Table Type		Question Type		Answer Type			Solution Type
					Domain	Format	Free-text	MC	Text	Integer	Decimal	
Dolphin18K (2016)	831	✗	✓	✗	✗	✗	✓	✗	✗	✓	✓	formula
DRAW-1K (2017)	1,000	✗	✓	✗	✗	✗	✓	✗	✗	✓	✓	formula
Math23K (2017)	23,162	✗	✓	✗	✗	✗	✓	✗	✗	✓	✓	formula
MathQA (2019)	37,297	✗	✓	✗	✗	✗	✗	✓	✗	✓	✓	formula
ASDiv (2020)	2,305	✗	✓	✗	✗	✗	✓	✗	✓	✓	✓	formula
SVAMP (2021)	1,000	✗	✓	✗	✗	✗	✓	✗	✗	✓	✗	formula
GSM8K (2021)	8,792	✗	✓	✗	✗	✗	✓	✗	✗	✓	✗	text
IconQA (2021b)	107,439	✗	✓	✗	✗	✗	✓	✓	✓	✓	✗	✗
FinQA (2021)	8,281	2,766	✓	76.6%	finance	text	✓	✗	✗	✓	✓	program
TAT-QA (2021)	16,552	2,747	50.0%	✓	finance	text	✓	✗	✗	✓	✓	✗
MultiHiertt (2022)	10,440	9,843	✓	89.8%	finance	text	✓	✗	✗	✓	✓	✗
TabMWP (ours)	**38,431**	**37,644**	✓	✓	**open**	**text***	✓	✓	✓	✓	✓	**text**

Dynamic Prompt Learning via Policy Gradient (PromptPG)

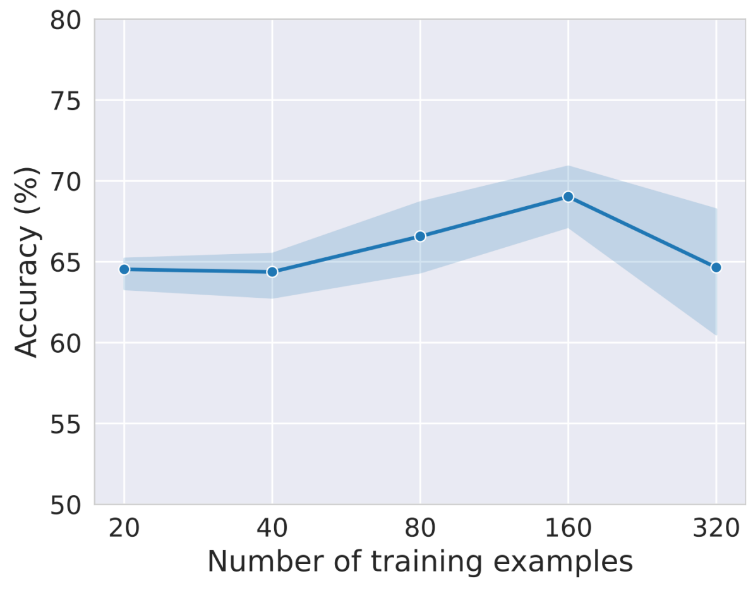
Algorithm 1 Dynamic Prompt Learning via Policy Gradient (PromptPG)

```
Input: Initial policy $\pi_{\theta_0}$, training example set $P_{\text{train}}$, candidate example set $E_{\text{cand}}$, # of training epochs $N$
Output: Learned policy $\pi_\theta$
function REINFORCE($\pi_{\theta_0}$, $P_{\text{train}}$, $E_{\text{cand}}$, $N$)
    Initialize policy network $\pi$ with parameter $\theta_0$
    for epoch = $1, 2, ..., N$ do
        for $P_{\text{batch}} \in P_{\text{train}}$ do                      ▷ get a batch from the training set
            $\mathcal{L}_{\text{batch}} \leftarrow 0$
            for $p_i \in P_{\text{batch}}$ do
                Sample $e_i^k \sim \pi_\theta(e_i|p_i), e_i^k \in E_{\text{cand}}, k = \{1, ..., K\}$    ▷ $K$ is # of in-context examples
                $\hat{a}_i \leftarrow$ GPT-3($e_i^1, ..., e_i^k, p_i$)        ▷ $\hat{a}_i$ is the GPT-3 generated answer
                $r_i \leftarrow$ EVAL($\hat{a}_i, a_i$), $r_i \in \{-1, 1\}$    ▷ $a_i$ is the ground truth answer of $p_i$
                $\mathcal{L}_{\text{batch}} \leftarrow \mathcal{L}_{\text{batch}} - r_i \cdot \ln \pi_\theta(e_i|p_i)$
            end for
            Optimize $\mathcal{L}_{\text{batch}}$ wrt. $\theta$
        end for
    end for
    return $\pi_\theta$
end function
```

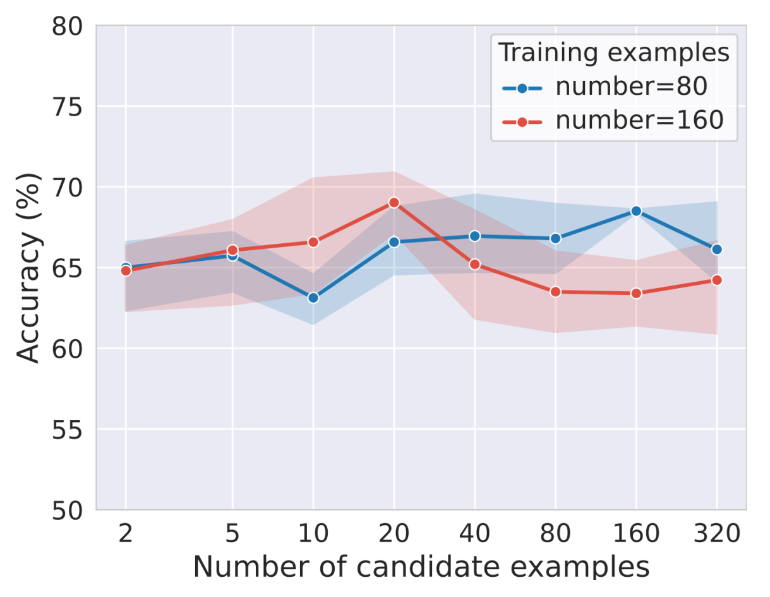
- Provided with a few **in-context examples**, GPT-3 can generate the answer for a test example
- This type of few-shot learning can be highly **unstable** across **different selections** of in-context examples
- It could be worse on TabMWP since problems are distributed across **diverse** question types and table layouts
- Our proposed **PromptPG** can **learn to select** in-context examples from candidates via **policy gradient**
- An agent learns to find optimal in-context examples from a candidate pool, with the goal of **maximizing the prediction rewards** on given training examples when interacting with the GPT-3 environment

Experimental Results on TabMWP

Method	Training Data	Selection Strategy	Question Types		Answer Types					Grades		Avg.
			FREE	MC	INT	DEC	EXTR	BOOL	OTH	1-6	7-8	
Heuristic Baselines												
Heuristic guess	-	-	6.71	39.81	8.37	0.26	30.80	51.22	26.67	17.55	12.27	15.29
Human performance	-	-	84.61	93.32	84.95	83.29	97.18	88.69	96.20	94.27	81.28	90.22
pre-trained Baselines												
UnifiedQA$_{\text{SMALL}}$	-	-	1.18	43.62	1.37	0.43	38.70	49.78	37.14	15.57	7.65	12.18
UnifiedQA$_{\text{BASE}}$	-	-	4.60	43.02	5.28	1.97	37.08	50.11	38.10	17.14	11.11	14.56
UnifiedQA$_{\text{LARGE}}$	-	-	4.48	48.80	5.19	1.72	48.33	50.33	40.00	19.78	10.87	15.96
TAPEX$_{\text{BASE}}$	-	-	7.32	39.76	8.68	2.06	35.06	47.11	20.95	18.67	11.81	15.73
TAPEX$_{\text{LARGE}}$	-	-	8.80	46.59	10.62	1.72	46.91	48.11	30.48	22.65	13.18	18.59
fine-tuned Baselines												
UnifiedQA$_{\text{SMALL}}$	23,059	-	22.27	51.31	27.27	2.83	52.28	48.11	69.52	35.85	21.71	29.79
UnifiedQA$_{\text{BASE}}$	23,059	-	34.02	70.68	40.74	7.90	84.09	55.67	73.33	53.31	30.46	43.52
UnifiedQA$_{\text{LARGE}}$	23,059	-	48.67	82.18	55.97	20.26	94.63	68.89	79.05	65.92	45.92	57.35
TAPEX$_{\text{BASE}}$	23,059	-	39.59	73.09	46.85	11.33	84.19	61.33	69.52	56.70	37.02	48.27
TAPEX$_{\text{LARGE}}$	23,059	-	51.00	80.02	59.92	16.31	**95.34**	64.00	73.33	67.11	47.07	58.52
Prompting Baselines w/ GPT-3												
Zero-shot	-	-	53.57	66.67	55.55	45.84	78.22	55.44	54.29	63.37	48.41	56.96
Zero-shot-CoT	-	-	54.36	66.92	55.82	48.67	78.82	55.67	51.43	63.62	49.59	57.61
Few-shot (2-shot)	2	Random	54.69	64.11	58.36	40.40	75.95	52.41	53.02	63.10	49.16	57.13
Few-shot-CoT (2-shot)	2	Random	60.76	69.09	60.04	63.58	76.49	61.19	**67.30**	68.62	55.31	62.92
PromptPG *w/ GPT-3 (Ours)*												
Few-shot-CoT (2-shot)	160+20	Dynamic	**66.17**	**74.11**	**64.12**	**74.16**	76.19	**72.81**	65.71	**71.20**	**64.27**	**68.23**$_{5.31\uparrow}$



Using **160 training examples** performs the best given 20 candidate examples.



Using **20 candidate examples** performs the best given 160 training examples.

Selection strategy	Acc. (%)
Same question type	66.2 ± 0.60
Same answer type	67.9 ± 0.38
Same grade level	67.9 ± 1.87
Most complex (# of table cells)	64.0 ± 0.42
Most complex (# of ques. words)	68.2 ± 0.26
Random selection	65.2 ± 4.01
Manual selection (fixed w/ top 2)	66.9 ± 0.00
Nearest neighbor	68.2 ± 0.29
PromptPG (Ours)	**70.9 ± 1.27**

Our **PromptPG** improve the accuracy with largely reduced randomness.

Case Study

▷ *In-context example (ID: 13974)*

Table:
heart-shaped beads | \$3/kilogram
rectangular beads | \$2/kilogram
spherical beads | \$2/kilogram
oval beads | \$2/kilogram

heart-shaped beads	\$3/kilogram
rectangular beads	\$2/kilogram
spherical beads	\$2/kilogram
oval beads	\$2/kilogram

Question: Rebecca bought 2.5 kilograms of oval beads. How much did she spend? (Unit: \$)
Answer:
(Step 1) Find the cost of the oval beads. Multiply the price per kilogram by the number of kilograms.
(Step 2) \$2 × 2.5 = \$5
(Step 3) She spent \$5. The answer is 5.

▷ *Test example (ID: 17417)*

Table:
[TITLE]: Birthday party
Activity | Parents | Children
Singing | 14 | 20
Eating cake | 5 | 10
Jumping rope | 16 | 20
Swimming | 16 | 19
Playing tag | 4 | 9

Birthday party		
Activity	**Parents**	**Children**
Singing	14	20
Eating cake	5	10
Jumping rope	16	20
Swimming	16	19
Playing tag	4	9

Question: At Josie's birthday party, children and parents celebrated by participating in some activities. How many more children than parents are playing tag? (Unit: children)
Answer:
(Step 1) **To find the difference between the number of children and parents playing tag, subtract the number of parents from the number of children.**
(Step 2) **9 - 4 = 5**
(Step 3) **There are 5 more children than parents playing tag. The answer is 5.**

Our **PromptPG** tends to select examples that have **multiple reasoning steps** and **similar abilities in mathematical reasoning**, which results in higher prediction accuracy.

▷ *In-context example (ID: 18129)*

Table:
[TITLE]: Children's weights (lbs)
Stem | Leaf
1 | 7
2 | 4
3 |
4 |
5 | 2, 2, 8
6 | 6
7 | 1, 3
8 | 7, 8
9 | 0

Children's weights (lbs)	
Stem	**Leaf**
1	7
2	4
3	
4	
5	2, 2, 8
6	6
7	1, 3
8	7, 8
9	0

Question: Dr. Thornton, a pediatrician, weighed all the children who recently visited her office. How many children weighed exactly 52 pounds? (Unit: children)
Answer:
(Step 1) For the number 52, the stem is 5, and the leaf is 2. Find the row where the stem is 5. In that row count all the leaves equal to 2.
(Step 2) You counted 2 leaves, which are blue in the stem-and-leaf plot above.
(Step 3) He will spend \$6.80. The answer is 6.80.

▷ *Test example (ID: 17417)*

Table:
[TITLE]: Birthday party
Activity | Parents | Children
Singing | 14 | 20
Eating cake | 5 | 10
Jumping rope | 16 | 20
Swimming | 16 | 19
Playing tag | 4 | 9

Birthday party		
Activity	**Parents**	**Children**
Singing	14	20
Eating cake	5	10
Jumping rope	16	20
Swimming	16	19
Playing tag	4	9

Question: At Josie's birthday party, children and parents celebrated by participating in some activities. How many more children than parents are playing tag? (Unit: children)
Answer:
(Step 1) For the activity "Playing tag", **the stem is 4, and the leaf is 9. Find the row where the stem is 4. In that row, count all the leaves equal to 9.**
(Step 2) **You counted 9 leaves, which are blue in the stem-and-leaf plot above. 9 children** are playing tag. **The answer is 9.**

Nearest neighbor search selects the **"superficially" similar** examples to the test example.