EXPLORA: Efficient Exemplar Subset Selection for Complex Reasoning

Kiran Purohit[1], Venktesh V[2], Raghuram Devalla[1], Krishna Mohan Yerragorla[1], Sourangshu Bhattacharya[1], Avishek Anand[2]

[1]CNeRG Lab, Indian Institute of Technology Kharagpur, [2]TU Delft

EMNLP 2024

The Big Question

- How Can We Tackle **Complex Reasoning** Across Text and Hybrid Sources?
- Can **In-Context Learning** Unlock the Power of Large Language Models?
- What's the Secret to Boosting ICL Performance? The **Right Exemplars!**

Prompting Ingredients for ICL

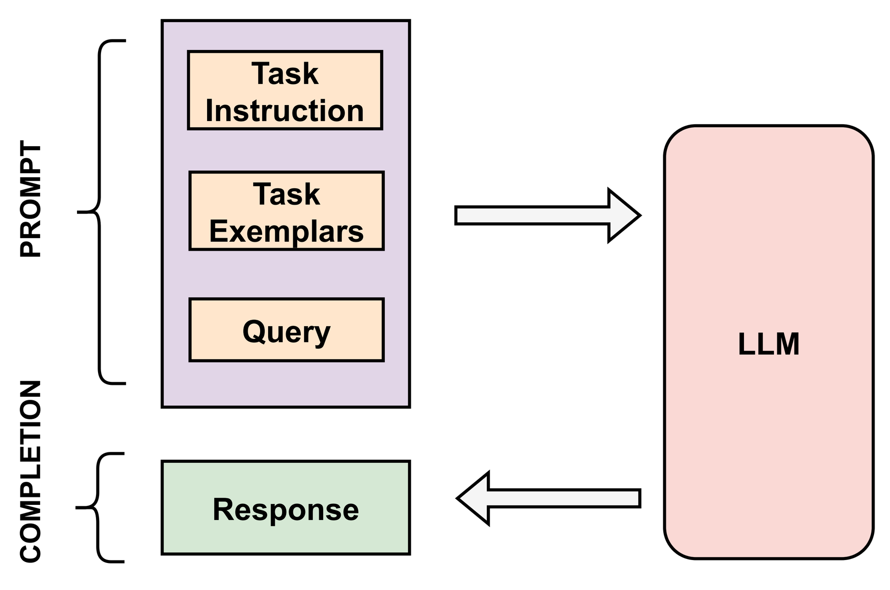

Figure 1: Block Diagram of ICL

LLM Calls and Runtime Comparison

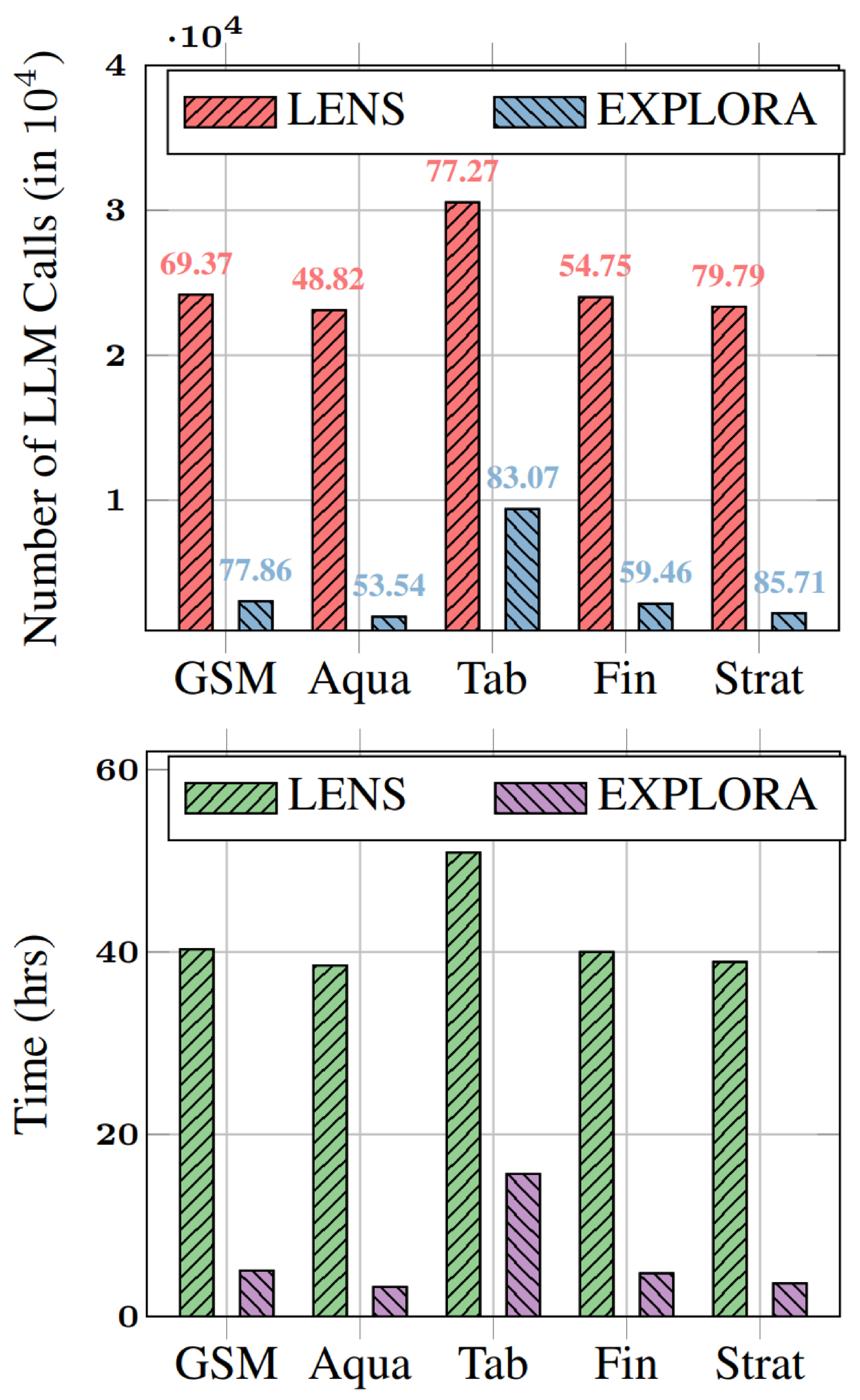

Figure 4: (Top) **LLM calls** LENS vs EXPLORA (y-axis) with corresponding EM scores indicated on top of bars. (Bottom) **Runtime** comparison LENS vs EXPLORA.

Key Contributions

✔ Proposed a novel top-l exemplar-subset selection approach, using a scoring function $\sigma(\alpha, .)$.

✔ Introduced a novel sampling-based bandit algo for efficiently learning the parameters of σ.

For Further Information

https://github.com/kiranpurohit/EXPLORA

Acknowledgements: ACM India-IARCS, CNeRG Travel Grant

Website

Twitter

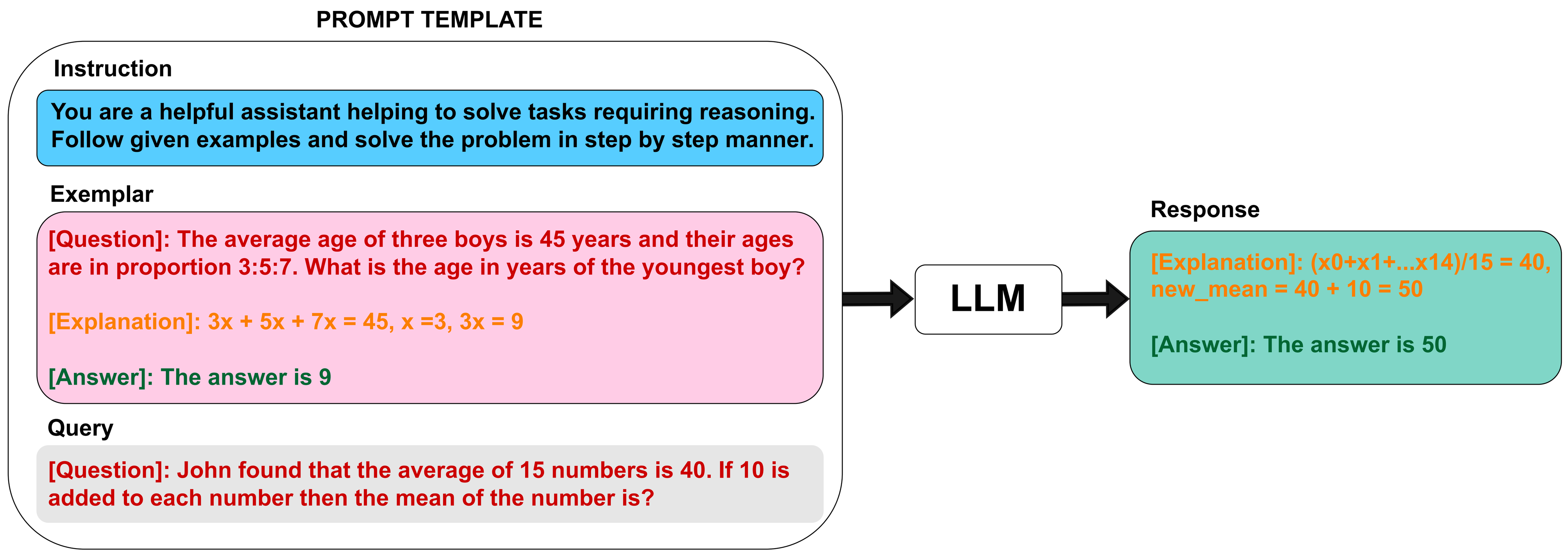

Figure 2: LLM is provided with a prompt that includes <Question-Explanation-Answer> triplet referred to as exemplars.

EXPLORA: Our Approach

Static exemplar subset selection method for black-box language models without access to its parameters.

❶ **Scoring Function** (σ): Linear function for approximating validation loss based on similarity features between exemplars and validation examples.

$$\sigma(\vec{\alpha}, S) = \frac{1}{m}\sum_{j=1}^{m}\sum_{i=1}^{n}\alpha_i \mathbb{1}(x_i \in S)E_{ij} \quad (1)$$

❷ **Sampling-Based Bandit Algorithm**: Efficiently learns parameters (α) of σ to identify top-l exemplar subsets. α is computed by minimizing the following loss function:

$$\mathcal{L}(\vec{\alpha}; U_t, V_t) = \sum_{S \in U_t}(L(S, \mathcal{V}) - \sigma(\vec{\alpha}, S))^2 + \sum_{S' \sim V_t}(L(S', \mathcal{V}) - \sigma(\vec{\alpha}, S'))^2 \quad (2)$$

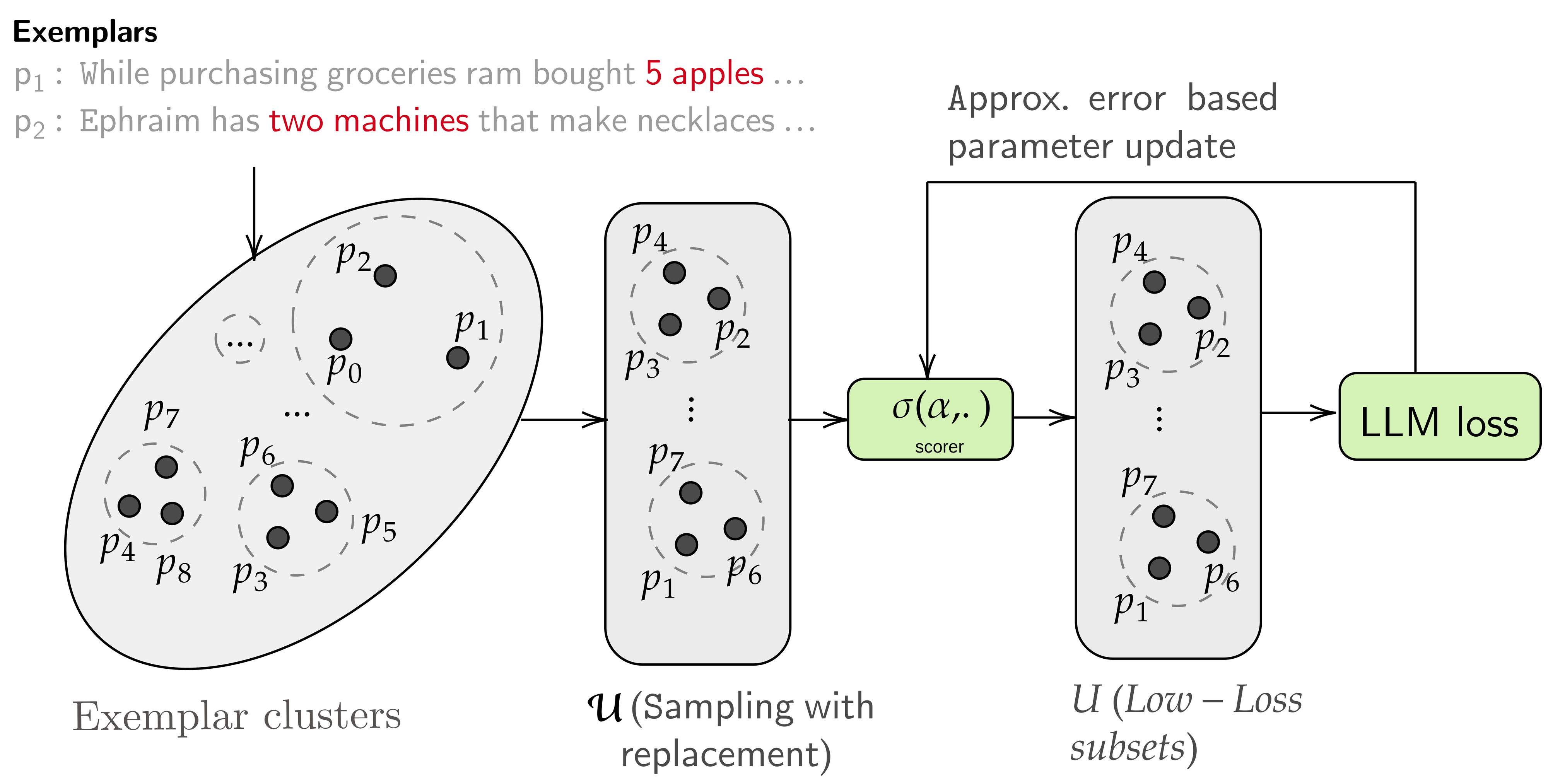

Figure 3: **Overview of EXPLORA:** In each round, parameters of scoring function $\sigma(\alpha, .)$ are computed by minimizing the loss function (2). σ guides the selection of the subset from $\mathcal{U} \setminus U$ with the lowest loss, which is then used to update U. This iterative updating process ensures U maintains low-loss subsets, leading to a more accurate estimation of α in subsequent rounds.

Performance Comparison

Method	GSM8K	AquaRat	TabMWP	FinQA	StrategyQA
dynamic					
KNN	53.45	51.96	77.07	51.52	81.83
MMR	54.36	51.18	77.32	49.87	82.86
static					
Manual Few-Shot COT	73.46	44.88	71.22	52.22	73.06
GraphCut	66.19	47.24	60.45	52.31	80.00
FacilityLocation	68.61	48.43	67.66	36.79	81.63
LENS	69.37	48.82	77.27	54.75	79.79
EXPLORA	77.86	53.54	83.07	59.46	85.71

Table 1: Results across datasets in transfer setting using gpt-3.5-turbo with exemplars selected from Mistral-7b.

Key Findings

- EXPLORA outperforms both static and dynamic exemplar selection methods.
- Exemplars selected using smaller LLMs transfer well to larger LLMs.
- Significantly reduces LLM calls to ∼11% of LENS and also reduces running time.
- Achieves substantial performance improvement of 12.24% to LENS