Augmenting Large Language Models with Symbolic Rule Learning for Robust Numerical Reasoning

Hadeel Al-Negheimish[1], Pranava Madhyastha[2,1], Alessandra Russo[1]

[1]Imperial College London, [2]City, University of London

Summary

- While LLM prompting has been used to elicit reasoning, **numerical reasoning in MRC** remains an open challenge.
- We propose a **neuro-symbolic approach** that only needs few-shot examples and evaluate it on different splits of the DROP benchmark.
- In addition to improving pure LLM performance, this approach provides **interpretable and verifiable** reasoning, maintaining **faithfulness to the passage**

Collecting few-shot examples

We build upon a small set of annotated examples[1]:

- **Random:** Randomly collect 3 examples
- **KNN:** Collect the 3 examples with questions most similar to the test question based on QQP sentence embeddings[2]
- **Gold-type:** Each type is associated with 3 canonical examples, assume that the type is known (gold)
- **Predict-type:** An LLM is used to predict the type, using its canonical examples

Rule learning to recompose partial answers

Given partial answers to simpler subquestions, this component learns the *reasoning* required to reach the final answer, using only few-shot examples.

We utilise symbolic rule learning using ILASP, a logic-based machine learning system that can induce rules given observed data combined with background knowledge, by defining a Learning from Noisy Answer Sets task:

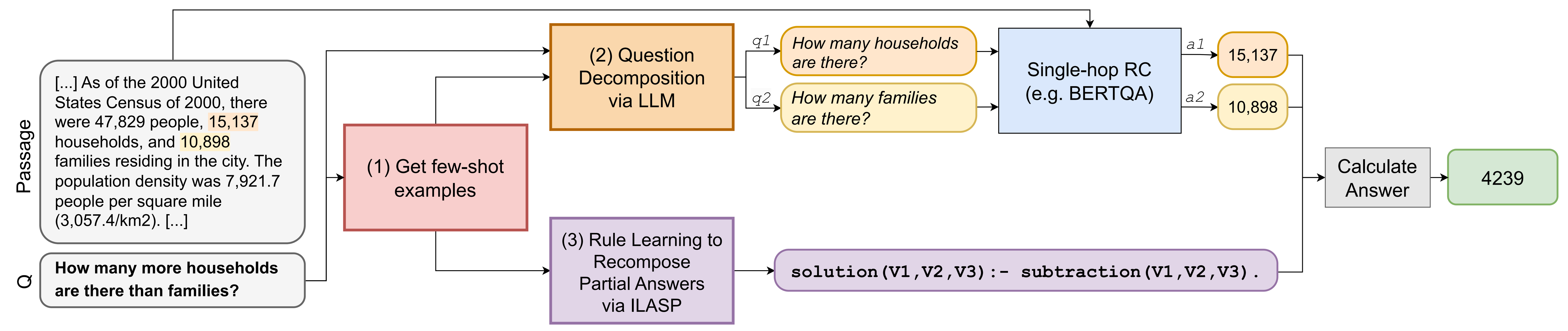
$$ILP_{LAS}^{noise} = \langle B, S_M, E \rangle$$

- B: encodes the background knowledge, we define the space of possible operations
- S_M: the hypothesis space, which is defined by mode declarations; which predicates can appear in the rules
- E: Examples with their associated penalties to allow robust learning with noise

Why symbolic learning?

- Allows learning from only a few examples
- Generalises beyond seen data
- Learned rule is interpretable

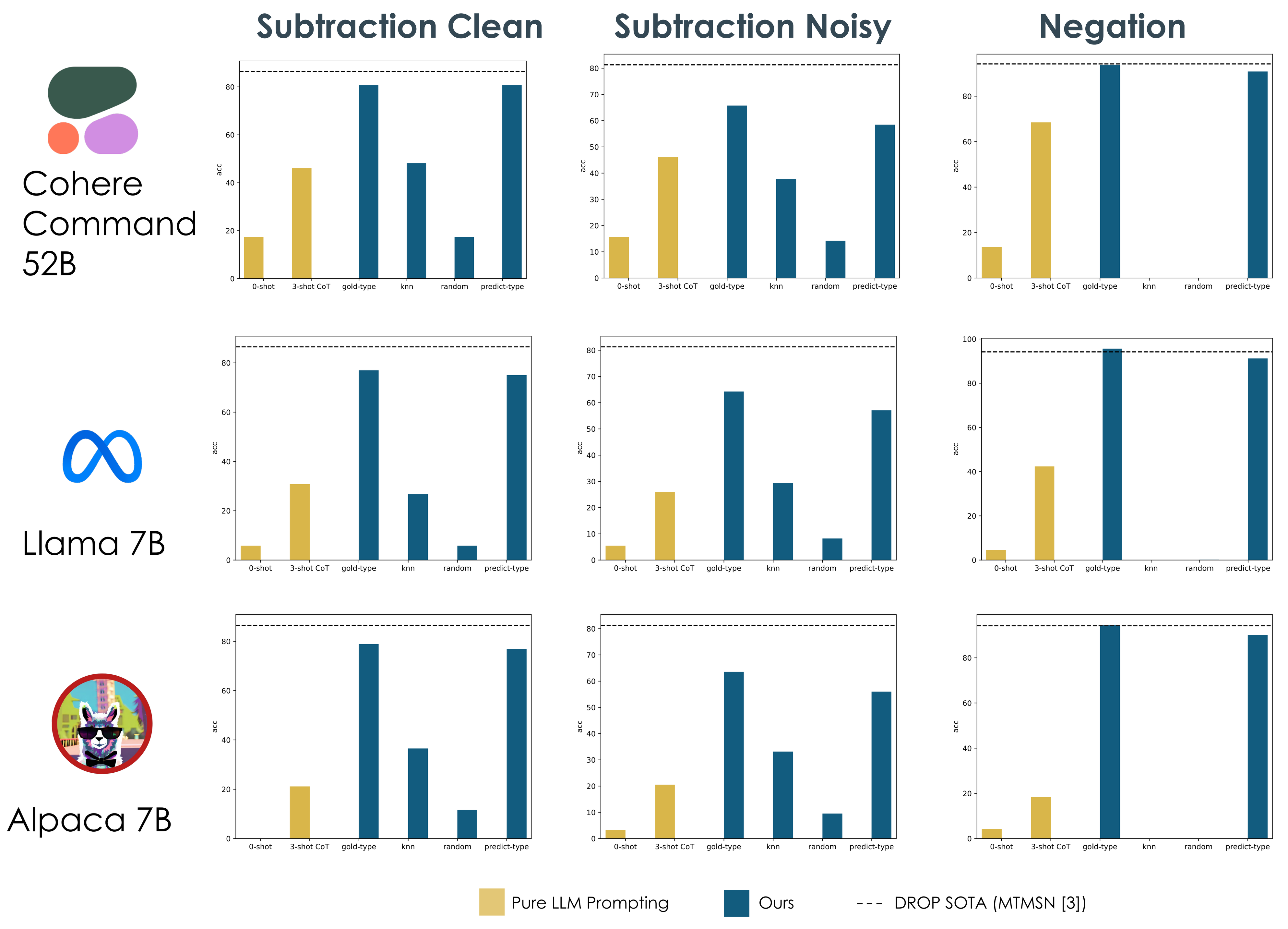
Approach Overview



Results

Key takeaways

- Our approach **performs competitively close** to DROP SOTA in many settings, **without needing data-intensive training**
- By using an RC model to answer partial questions, we force models to be faithful to the given context
- Type-prediction performs almost as well as gold-type for all models
- Our approach **improves upon all pure LLM prompting,** 0-shot and 3-shot Chain-of-Thought
- Our approach **bridges the performance gap** between small and large LLMs by simplifying the problem



[1] Dheeru Dua, Shivanshu Gupta, Sameer Singh, and Matt Gardner. Successive prompting for decomposing complex questions. EMNLP 2022. https://aclanthology.org/2022.emnlp-main.81
[2] Nils Reimers and Iryna Gurevych. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. EMNLP-IJCNLP 2019. https://aclanthology.org/D19-1410
[3] Minghao Hu, Yuxing Peng, Zhen Huang, and Dongsheng Li. A multi-type multi-span network for reading comprehension that requires discrete reasoning. EMNLP-IJCNLP 2019. https://aclanthology.org/D19-1170