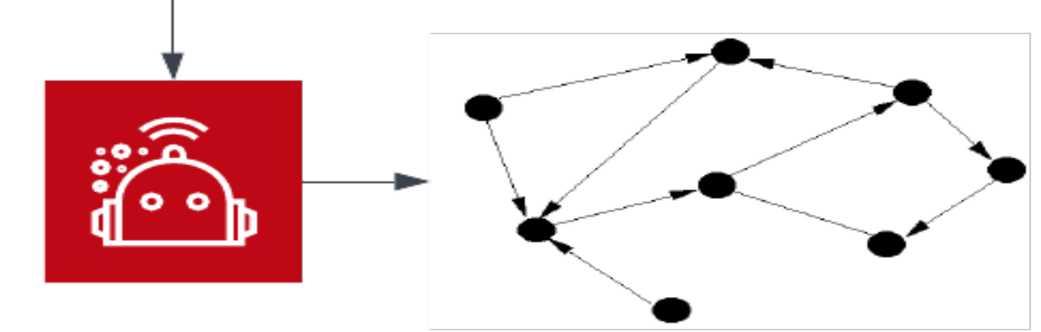
English to Linear Temporal Logic Translation using Neural Language Models

Mohamed Aghzal, Ziyu Yao, Erion Plaku
Department of Computer Science, George Mason University

Robot Motion Planning Overview



Introduction

In this work, we aim to synthesize natural language descriptions into linear temporal logic (LTL) specifications using neural language models.

Data Synthesis

We used following classes of primitive rules to synthesize the training data:

- **Eventually**
- **Order**
- **Arithmetic**
- **Redirection**

Compositional Generalization

To evaluate the model's compositional generalization capabilities, we designed our test set by drawing examples consisting of combinations of the primitive rules used above.

- ***In-Domain Generalization*** (1000)
- ***Combinations of rules*** (2750)
- ***Compositions of the same rule***(1500)

Prefix, Postfix, Infix

The generated LTL formulas can quickly grow in size and complexity

→ parsing such formulas "as they are" can be very hard

→ many parsers convert the formulas into a prefix or postfix representation in order to simplify parsing

Infix: (! p11 U (p10 & p14 & p12) & F (p11))
Prefix: & U ! p11 & & p10 p14 p12 F p11
Postfix: p11 ! p10 p14 & p12 & U p11 F &

Training

We trained a **BART** [1] using the original parameters introduced in their paper for 10 epochs. Training was done using two NVIDIA A100 80gb GPUs.
We trained separate models for postfix, prefix and infix representations for LTL.

Evaluation Metrics

Jaccard Similarity:

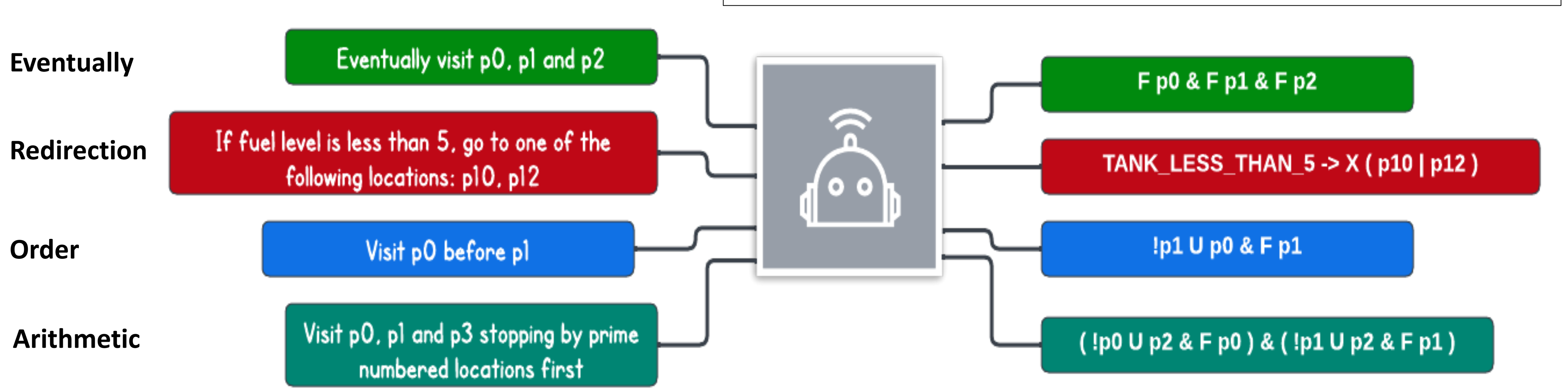
$$\frac{|A \cap B|}{|A \cup B|}$$

Dice Similarity:

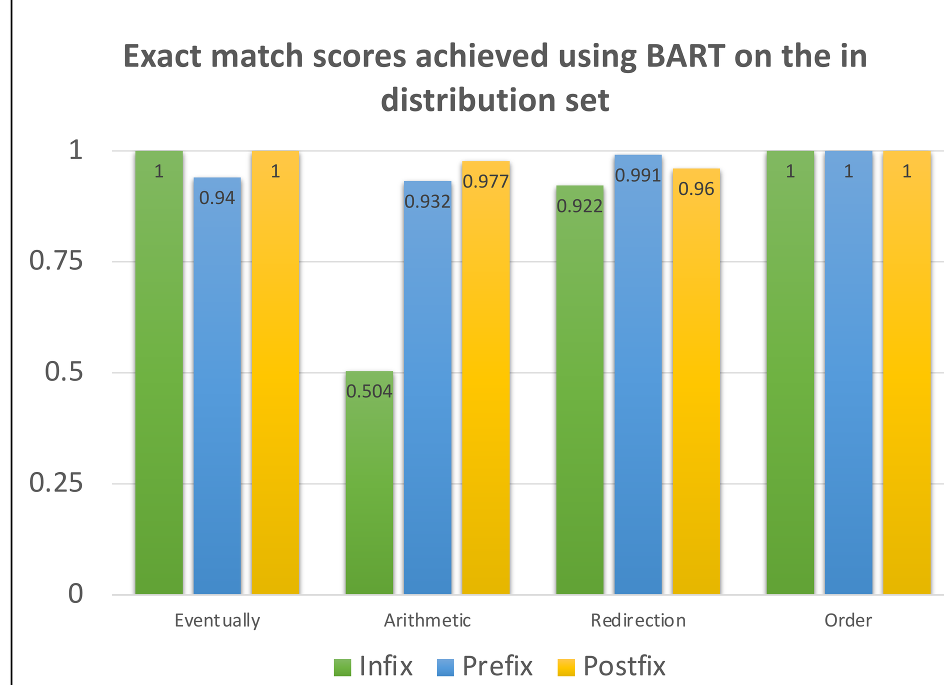
$$2 * \frac{|A \cap B|}{(|A| + |B|)}$$

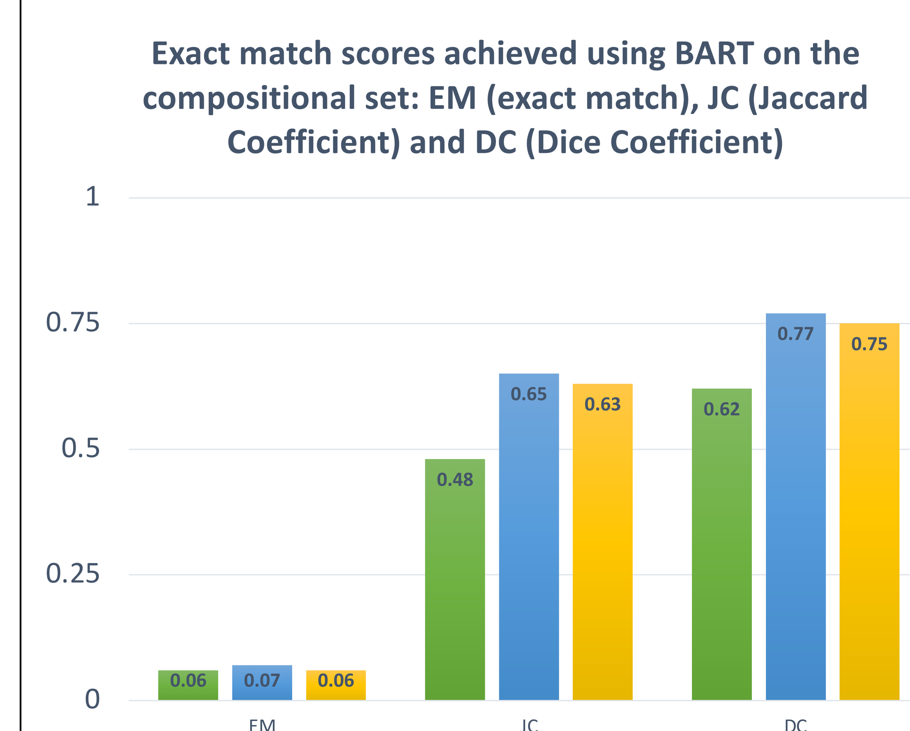
Where,

A: the set of all subformulas in the predicted formula
B: the set of all subformulas in the correct formula



Results





Error Analysis

Predicted: (!p14 U (p5 & p10 & p15 & p2) & F (p14))
Correct: (! p15 U (p5 & p10 & p14 & p2) & F (p15))
English: Visit p5, p10, p14, p15 and p2 stopping by multiples of 3 last.

Predicted: (! p7 U (p4 & p10 & p12 & p12) & F (p7))
Correct: (! p7 U (p4 & p10 & p12) & F (p7))
English: Visit p4, p10, p12 and p7 stopping by even numbered locations first.

Observations

- Using a postfix representation achieved the best performance with the prefix representation not falling too far behind
- The infix representation appears to struggle particularly with examples involving arithmetic
- The model does not appear to be able to generalize compositionally

Future Work

- Evaluate the expressions generated on a real motion planning environment
- Experiment with other language models, such as LLaMa or GPT-3
- Add a human feedback loop.

References

[1] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, & Luke Zettlemoyer (2019). BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension.