

Research Report: Hybrid Discrete-Transformer Algorithm for Synthetic PolyRule Reasoning

Agent Laboratory

June 25, 2025

Abstract

We propose a hybrid discrete-Transformer algorithm for Synthetic PolyRule Reasoning (SPR) that integrates discrete token extraction with a lightweight Transformer encoder to learn and verify abstract symbolic rules from sequences $S = [s_1, s_2, \dots, s_L]$ where each token combines a shape from $\{\triangle, \bullet, \circ\}$ and a color from $\{r, g, b, y\}$; the task is challenging because it requires disentangled representation learning of composite symbols and precise modeling of predicate-level dependencies such as shape-count, color-position, parity, and order, which are not easily captured by conventional continuous models. To address these challenges, our method employs a two-stage training protocol: in Stage 1, a discrete tokenization module is pre-trained using a JEPA-inspired auxiliary contrastive loss defined as $\mathcal{L}_{aux} = \|Drop(f(x)) - Drop(f(y))\|_2^2$ which converges to an average loss of approximately 0.0114, ensuring robust and disentangled feature representations; in Stage 2, the Transformer encoder, enhanced with positional encoding, and a predicate reasoning head are jointly fine-tuned using a composite loss $\mathcal{L} = \mathcal{L}_{cls} + 0.5 \mathcal{L}_{aux}$ that yields a joint loss of 0.7113. Comprehensive experiments on the SPR_BENCH dataset demonstrate a Development Set Shape-Weighted Accuracy (SWA) of 45.59% and a Final Test SWA

of 48.31%, as shown in the summary table

Metric	Value
<i>DevSWA</i>	45.59%
<i>TestSWA</i>	48.31%

; although these results do not yet match the state-of-the-art baselines of approximately 60.0% to 65.0%, they substantiate the potential of our hybrid framework, underscore the inherent difficulties in achieving systematic symbolic pattern recognition, and motivate future enhancements including extended training iterations, refined predicate integration strategies, and alternative augmentation methodologies.

1 Introduction

[INTRODUCTION HERE]

2 Background

[BACKGROUND HERE]

3 Related Work

[RELATED WORK HERE]

4 Methods

[METHODS HERE]

5 Experimental Setup

[EXPERIMENTAL SETUP HERE]

6 Results

[RESULTS HERE]

7 Discussion

[DISCUSSION HERE]