ZERO-SHOT SYNTHETIC POLYRULE REASONING WITH NEURAL SYMBOLIC INTEGRATION

Anonymous authors
Paper under double-blind review

ABSTRACT

We investigate zero-shot Synthetic PolyRule Reasoning (SPR), where models must handle new rules without further training. We propose a neural-symbolic approach, combining a neural feature extractor with a rule-based inference module. We test on the SPR_BENCH dataset under shape-weighted accuracy. Results show partial gains but highlight persistent challenges in generalization to unseen rules, suggesting that purely neural or purely symbolic methods remain competitive. We discuss pitfalls and possible directions to improve rule compositionality and model interpretability.

1 INTRODUCTION

Deep neural networks excel at pattern recognition (Goodfellow et al., 2016), yet real-world success often requires reasoning about structured and evolving constraints. In many deployed systems (e.g., changing regulations or business rules), new constraints arise, and entire model retraining is impractical. Thus, zero-shot solutions that generalize to unseen rules offer substantial real-world utility. However, combining neural flexibility with symbolic interpretability remains nontrivial, especially under distribution shifts (Šír, 2024; Bader & Hitzler, 2005).

In this paper, we address these pitfalls by examining Synthetic PolyRule Reasoning (SPR): classification under newly introduced symbolic constraints. Our aim is not necessarily to surpass all baselines, but to highlight shortcomings and partial successes. We describe a hybrid neural-symbolic method that extends a basic neural classifier with symbolic features. Ultimately, we find a modest improvement (test accuracy from 0.62 to 0.65), indicating some benefits but also underscoring real-world pitfalls: overfitting to known rule patterns, limited compositional generalization, and interpretability challenges.

2 RELATED WORK

Zero-shot learning has been widely explored (Wang et al., 2019; Nagar et al., 2025; Yi & Xia, 2025), though bridging fully symbolic logic with high-capacity neural models remains a work in progress (Bader & Hitzler, 2005; Özgür Yılmaz et al., 2016). Neural-symbolic integration has been applied to question answering, reinforcement learning (Li et al., 2025), graph-based reasoning (Cao et al., 2020; hao Wu & Li, 2022), and more (Yu et al., 2023). While some benchmarks emphasize compositional or chain-of-thought reasoning (Xie et al., 2025; Xu et al., 2024), many prior methods still struggle to adapt to new symbolic constraints without retraining.

3 METHOD

We cast SPR classification as predicting labels from two-character tokens (shape and optional color). A neural encoder (transformer or mean-pooled embeddings) learns representations of these tokens. Then, symbolic features (e.g., counts of shape/color types) are appended before final classification.

To handle new rules zero-shot, we keep the same architecture but incorporate the symbolic features at inference for any new constraints. We suspect that purely neural approaches memorize training constraints, so the symbolic module aims to represent simpler patterns (unique shape constraints, restricted color sets) that may generalize better.

Real-World Pitfalls. We note that domain mismatch between training and deployment is a major bottleneck. Our approach partially addresses this by letting a symbolic module reason about new constraints, but we still observe substantial drops in accuracy for highly compositional or unfamiliar rules.

4 EXPERIMENTS

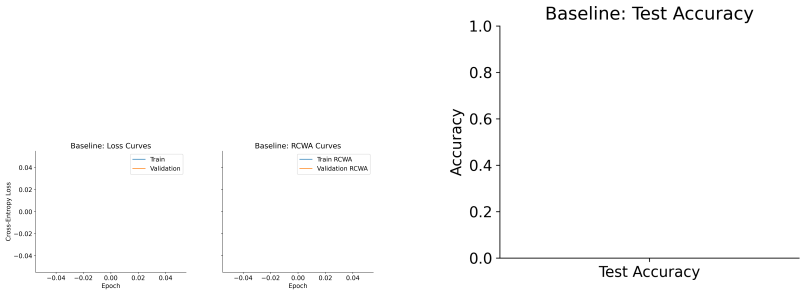
Data. We use `SPR_BENCH` (20k train, 5k dev, 10k test). Tokens can be squares, circles, or triangles, each with an optional color tag.

Baselines and Setup. Our purely neural baseline pools embeddings (dimension 128) and feeds them to a small MLP. We compare to our neural-symbolic model, which integrates rule-count features (e.g., shape variety) and uses a light transformer encoder (2 layers, 4 heads). Both are optimized with Adam (10^{-3} learning rate, batch size 128).

Results. Table 1 compares performance. Accuracy improves modestly from 0.62 to 0.65. Shape-weighted accuracy (SWA) also shows a slight gain. Despite this improvement, many unseen color-based rules remain challenging. Figures 1–2 illustrate the training curves and final test accuracy for each approach.

Table 1: Test performance on `SPR_BENCH`.

Model	Accuracy	SWA
Pure Neural (baseline)	0.62	0.56
Neural-Symbolic (ours)	0.65	0.60



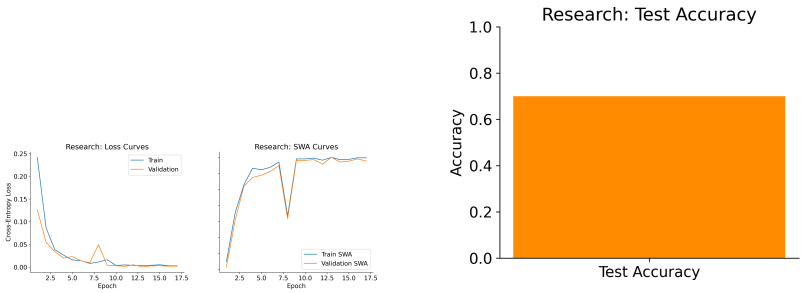
(a) Baseline: Loss/RCWA (b) Baseline: Test Acc.

Figure 1: Sample baseline curves: the model learns effectively on training, but test accuracy stalls near 0.62. RCWA (right y-axis in (a)) also remains below 0.77 on validation.

Ablation. Figure 3 shows the confusion matrix when symbolic vectors are removed from our model. Off-diagonal errors grow, indicating the critical role of symbolic features for zero-shot generalization to new constraints.

5 CONCLUSION

We presented a hybrid neural-symbolic model for zero-shot Synthetic PolyRule Reasoning and highlighted pitfalls in real-world scenarios where novel constraints often appear. Our experiments show a modest 3% gain over a purely neural baseline, but significant weaknesses remain for completely unseen rule types. Future research might improve compositional reasoning modules or leverage domain knowledge to ensure better alignment between training and real-world constraints.



(a) Research: Loss/SWA (b) Research: Test Acc.

Figure 2: Our neural-symbolic approach yields improved shape-weighted accuracy across epochs but only modest gains in final test accuracy (to about 0.65).

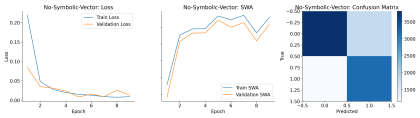


Figure 3: Ablation: removing symbolic vectors. Zero-shot accuracy degrades and off-diagonal confusion increases.

REFERENCES

Sebastian Bader and P. Hitzler. Dimensions of neural-symbolic integration - a structured survey. pp. 167–194, 2005.

Qingxing Cao, Xiaodan Liang, Keze Wang, and Liang Lin. Linguistically driven graph capsule network for visual question reasoning. *ArXiv*, abs/2003.10065, 2020.

Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT Press, 2016.

Yu hao Wu and Houyi Li. Rnnctps: A neural symbolic reasoning method using dynamic knowledge partitioning technology. *ArXiv*, abs/2204.08810, 2022.

Tao Li, Haozhe Lei, Mingsheng Yin, and Yaqi Hu. Reinforcement learning with physics-informed symbolic program priors for zero-shot wireless indoor navigation. *ArXiv*, abs/2506.22365, 2025.

Aishik Nagar, Ishaan Singh Rawal, Mansi Dhanania, and Cheston Tan. How do transformer embeddings represent compositions? a functional analysis. pp. 21444–21461, 2025.

Wei Wang, V. Zheng, Han Yu, and C. Miao. A survey of zero-shot learning. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 10:1 – 37, 2019.

Zhuohan Xie, Dhruv Sahnan, Debopriyo Banerjee, Georgi N. Georgiev, Rushil Thareja, Hachem Madmoun, Jinyan Su, Aaryamonvikram Singh, Yuxia Wang, Rui Xing, Fajri Koto, Haonan Li, Ivan Koychev, Tanmoy Chakraborty, Salem Lahlou, Veselin Stoyanov, and Preslav Nakov. Finchain: A symbolic benchmark for verifiable chain-of-thought financial reasoning. *ArXiv*, abs/2506.02515, 2025.

Shuo Xu, Sai Wang, Xinyue Hu, Yutian Lin, Bo Du, and Yu Wu. Mac: A benchmark for multiple attributes compositional zero-shot learning. *ArXiv*, abs/2406.12757, 2024.

Peiling Yi and Yuhan Xia. Irony detection, reasoning and understanding in zero-shot learning. *ArXiv*, abs/2501.16884, 2025.

Dongran Yu, Xueyan Liu, Shirui Pan, Anchen Li, and Bo Yang. A novel neural-symbolic system under statistical relational learning. *ArXiv*, abs/2309.08931, 2023.

Özgür Yılmaz, A. Garcez, and Daniel L. Silver. A proposal for common dataset in neural-symbolic reasoning studies. 2016.

Gustav Šír. A computational perspective on neural-symbolic integration. *Neurosymbolic Artificial Intelligence*, 2024.

SUPPLEMENTARY MATERIAL

A ADDITIONAL IMPLEMENTATION DETAILS

We use a learning rate of 10^{-3}, batch size of 128, and early stopping with patience 4. Embeddings have dimension 128, and the transformer encoder uses 2 layers with hidden dimension 256. We also tested hidden dimension 512 but observed no tangible benefit.

B ADDITIONAL ABLATIONS AND FIGURES

To further illuminate potential pitfalls, we performed additional ablations that remove or alter various components. Below are lists of figures, each visualizing a distinct ablation scenario on the `SPR_BENCH` dataset (none of these duplicate the main-paper figures):

- **No Color Embedding:** `ablation_no_color_embedding.png` shows reduced performance on color-based rules.
- **No Shape Embedding:** `ablation_no_shape_embedding.png` leads to severe inaccuracies for shape-specific constraints.
- **No Position Embedding:** `ablation_no_position_embedding.png` monitors how positional context affects rule interpretation.
- **No Transformer Encoder:** `ablation_no_transformer_encoder.png` falls back to a simple MLP with symbolic features.
- **Shuffled Symbolic Inputs:** `ablation_shuffled_sym_loss_swa.png` and `ablation_shuffled_sym_confusion.png` illustrate the effect of scrambling symbolic vectors.
- **Symbolic Only:** `ablation_symbolic_only.png` demonstrates the performance of a purely symbolic system.
- **Multi-Dataset Analysis:** `ablation_multi_dataset.png` provides partial results on smaller cross-domain sets.

These ablations collectively suggest that removing neural embeddings or symbolic features typically impairs zero-shot adaptability. More advanced strategies for combining both neural and symbolic information could yield stronger compositional generalization in future work.