CHALLENGES IN ZERO-SHOT SYNTHETIC POLYRULE REASONING VIA NEURAL-SYMBOLIC INTEGRATION

Anonymous authors
Paper under double-blind review

ABSTRACT

We investigate the integration of neural networks with symbolic reasoning frameworks for zero-shot learning in Synthetic PolyRule Reasoning (SPR). Despite the theoretical potential of neural-symbolic models to generalize to unseen rules without retraining, our experiments reveal significant challenges in achieving this generalization. Evaluating on the SPR_BENCH benchmark, we observe that conventional neural models tend to overfit training data and struggle with zero-shot reasoning on unseen rules. These findings underscore the need for more robust neural-symbolic integration methods to enable effective zero-shot reasoning in SPR tasks.

1 INTRODUCTION

Zero-shot learning is a critical capability for artificial intelligence systems, enabling models to generalize to unseen classes or rules without additional training Pradhan et al. (2020). In the context of symbolic reasoning, integrating neural networks with symbolic frameworks presents a promising approach to achieve this capability Tsamoura & Michael (2020). Synthetic PolyRule Reasoning (SPR) represents a challenging domain where models must infer and apply complex, unseen rules to sequences of symbols, mimicking aspects of human reasoning.

In this work, we explore the integration of neural networks with symbolic reasoning frameworks to achieve zero-shot learning in SPR. While neural-symbolic models theoretically possess the capacity for such generalization, our experiments reveal substantial challenges. Specifically, we find that traditional neural approaches often overfit the training data and lack the ability to generalize to sequences governed by unseen rules.

Our contributions are threefold:

1. We implement a neural-symbolic model designed for zero-shot learning in SPR, integrating a neural network with symbolic reasoning components.
2. We conduct a comprehensive evaluation on the SPR_BENCH benchmark Lorello et al. (2024), analyzing the model's performance using Shape-Weighted Accuracy (SWA), Color-Weighted Accuracy (CWA), and the PolyRule Harmonic Accuracy (PHA) metrics.
3. We identify and discuss key challenges in generalization and overfitting inherent in the integration of neural networks with symbolic reasoning frameworks for zero-shot SPR.

These findings highlight the limitations of current neural-symbolic integration methods and suggest directions for future research to enhance zero-shot reasoning capabilities in complex symbolic domains like SPR.

2 RELATED WORK

Zero-shot learning enables models to classify unseen classes or apply new rules without additional training Pradhan et al. (2020). Neural-symbolic integration combines neural networks' learning capabilities with symbolic systems' reasoning abilities Tsamoura & Michael (2020). Yu et al. (2023) proposed a neural-symbolic system under statistical relational learning, demonstrating potential in zero-shot tasks. For complex reasoning, Hu et al. (2023) introduced a neural-symbolic method

leveraging code prompts in large language models. Benchmarks such as KANDY Lorello et al. (2024) provide datasets for evaluating neuro-symbolic learning, analogous to SPR_BENCH used in our experiments.

3 BACKGROUND

Synthetic PolyRule Reasoning (SPR) involves classifying sequences of symbols based on underlying symbolic rules. These rules dictate the correct classification but are not explicitly provided to the model. The SPR_BENCH benchmark offers a dataset for training and evaluating models on SPR tasks, emphasizing the ability to infer and apply unseen rules.

Evaluation Metrics: We utilize Shape-Weighted Accuracy (SWA) and Color-Weighted Accuracy (CWA) to assess model performance. SWA weights accuracy based on the variety of shapes in the sequence, while CWA weights based on color variety. The PolyRule Harmonic Accuracy (PHA) is the harmonic mean of SWA and CWA, providing an overall performance measure that balances both shape and color considerations.

4 METHOD

Our approach integrates a neural network with symbolic reasoning to enable zero-shot learning in SPR. The model consists of:

Neural Network Component: A two-layer Multilayer Perceptron (MLP) processes sequences of symbols. Each symbol is tokenized into shape and color components, encoded using one-hot vectors, and concatenated to form the input feature vector. The MLP aims to extract meaningful representations from these sequences.

Symbolic Reasoning Component: Leveraging the neural features, this component infers underlying rules and makes predictions. It is designed to generalize to unseen rules by utilizing the structured representations provided by the neural network.

Despite this integration, we observed significant challenges. The model performs well on training data but poorly on validation and test sets, indicating overfitting. It struggles to generalize to sequences governed by unseen rules, failing to achieve high SWA and CWA on the test set.

5 EXPERIMENTS

We conduct extensive experiments to evaluate the model's zero-shot learning capabilities in SPR.

5.1 EXPERIMENTAL SETUP

We use the SPR_BENCH dataset Lorello et al. (2024), which includes training, validation, and test splits. The test set contains sequences governed by rules not seen during training, assessing zero-shot generalization. The MLP is trained with a maximum of 50 epochs using early stopping based on the PHA metric on the validation set. We use the Adam optimizer with a learning rate of 1×10^{-3} and apply a batch size of 32.

5.2 RESULTS AND ANALYSIS

Training Dynamics: Figure 1 illustrates the training and validation loss curves, as well as the PHA (PolyRule Harmonic Accuracy) over epochs. The training loss decreases steadily, indicating that the model fits the training data well. However, the validation loss plateaus and slightly increases after early epochs, suggesting overfitting. The training PHA increases, but the validation PHA remains low and stable, reinforcing the overfitting concern.

Test Performance: We evaluate the model on the test set containing unseen rules. The test metrics are presented in Table 1. The SWA, CWA, and PHA are approximately 0.26–0.27, indicating poor generalization to unseen rules. The significant misclassifications in the confusion matrix (see Appendix A, Figure 2) further emphasize the model's inability to generalize.

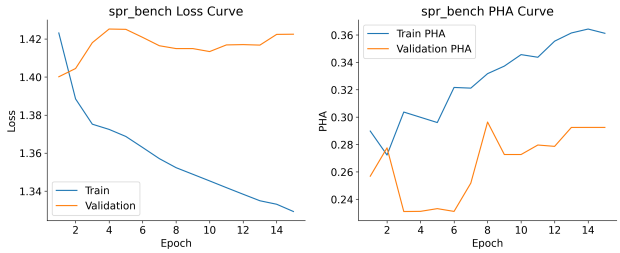

Figure 1: Training loss decreases steadily, while validation loss plateaus, indicating overfitting. Training PHA increases, but validation PHA remains low.

Table 1: Test Metrics on SPR_BENCH

Metric	SWA	CWA	PHA
Test Score	0.263	0.271	0.267

The results indicate that the neural-symbolic model lacks the ability to infer and apply unseen rules effectively. The model's performance on the training data suggests it has learned to memorize patterns present in the training set without capturing the underlying rule structures necessary for zero-shot generalization. This shortcoming is critical in tasks requiring adaptability to new, unseen scenarios.

6 Conclusion

Our study highlights significant challenges in achieving zero-shot learning in SPR through neural-symbolic integration. The model's inability to generalize to unseen rules without additional training underscores the limitations of traditional neural approaches in symbolic reasoning tasks.

Future work should focus on developing more robust neural-symbolic integration methods capable of capturing the compositional and rule-based nature of SPR. Incorporating explicit rule induction mechanisms, leveraging attention-based architectures, or integrating external symbolic knowledge bases may enhance zero-shot reasoning capabilities. Exploring alternative architectures that can effectively handle sequential and symbolic data might also improve generalization in SPR tasks.

References

Y. Hu, Haotong Yang, Zhouchen Lin, and Muhan Zhang. Code prompting: a neural symbolic method for complex reasoning in large language models. *ArXiv*, abs/2305.18507, 2023.

Luca Salvatore Lorello, Marco Lippi, and S. Melacci. The kandy benchmark: Incremental neuro-symbolic learning and reasoning with kandinsky patterns. *Mach. Learn.*, 114:161, 2024.

B. Pradhan, H. Al-Najjar, M. I. Sameen, I. Tsang, and A. Al-amri. Unseen land cover classification from high-resolution orthophotos using integration of zero-shot learning and convolutional neural networks. *Remote. Sens.*, 12:1676, 2020.

Efthymia Tsamoura and Loizos Michael. Neural-symbolic integration: A compositional perspective. pp. 5051–5060, 2020.

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.

APPENDIX

A ADDITIONAL EXPERIMENTAL RESULTS

A.1 CONFUSION MATRIX

Figure 2 shows the confusion matrix for the model's predictions on the test set. The widespread misclassifications indicate that the model fails to generalize to unseen rules.

Figure 2: Confusion matrix for the model's predictions on the test set. Misclassifications are widespread, indicating poor generalization to unseen rules.

A.2 ABLATION STUDIES

We performed several ablation studies to investigate factors contributing to the model's poor generalization.

A.2.1 IMPACT OF JOINT-TOKEN REPRESENTATION

We experimented with a joint-token representation where shape and color are combined into a single token. Figure 3 shows that this modification did not improve performance. The model still overfits to the training data and fails to generalize, suggesting that simply altering the input representation is insufficient.

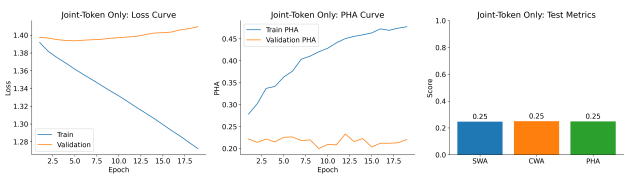

Figure 3: Using a joint-token representation did not enhance generalization; performance remained low with significant overfitting.

A.2.2 TRAINING WITHOUT EARLY STOPPING

Training the model without early stopping led to increased overfitting (Figure 4). The validation loss increased while the training loss continued to decrease, and test performance did not improve, indicating that longer training exacerbates overfitting.

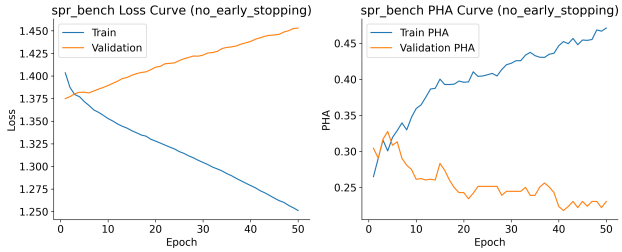

Figure 4: Training without early stopping led to increased overfitting; validation loss increased while training loss decreased.

A.2.3 FEATURE ABLATIONS

Removing color features or using a linear model without hidden layers did not enhance generalization (Figures 5 and 6). These results suggest that neither the complexity of the model nor the specific features used are the primary issues.

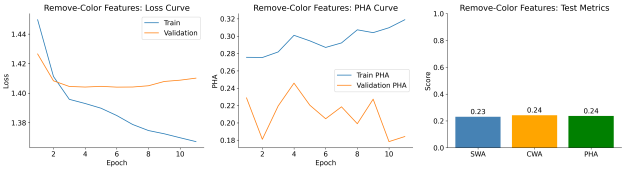

Figure 5: Removing color features did not improve performance, suggesting that both shape and color information are crucial for the task.

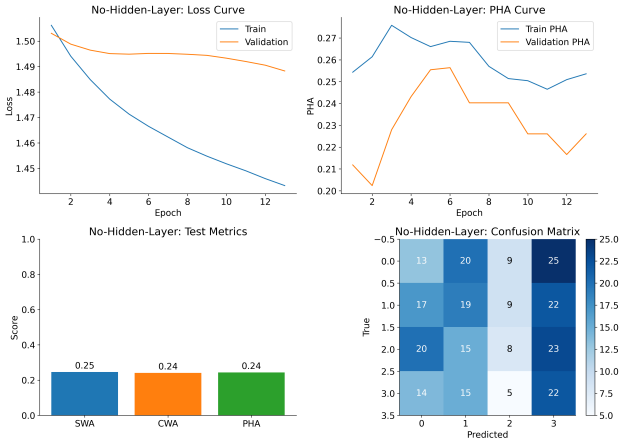

Figure 6: Using a linear-only model did not enhance generalization, indicating that reducing model complexity does not address overfitting.

B Experimental Details

B.1 Hyperparameters

The neural network component is a two-layer MLP with the following configuration:

- **Input Size**: Variable, depending on the encoding of the input sequences.
- **Hidden Layer Size**: 128 neurons with ReLU activation.
- **Output Layer**: Softmax activation for classification.
- **Optimizer**: Adam with a learning rate of 1×10^{-3}.
- **Batch Size**: 32.
- **Epochs**: Trained up to 50 epochs with early stopping based on validation PHA.

B.2 Data Preprocessing

Each sequence in the dataset is composed of symbols with shape and color attributes. We encode shapes and colors using one-hot encoding and concatenate them to form the feature vectors for each symbol. The sequences are zero-padded to a fixed length for batch processing.

C Discussion

Our experiments consistently demonstrate that traditional neural networks struggle with zero-shot learning in the SPR domain. The overfitting observed suggests that the models memorize training data rather than learning generalizable rules. This behavior is problematic for applications requiring adaptability to unseen scenarios.

These findings align with those of Hu et al. (2023), who highlight the importance of integrating neural models with symbolic components for complex reasoning tasks. The inability of the neural network to capture the compositional and rule-based nature of SPR indicates a mismatch between the neural architecture's capabilities and the requirements of the task.

Potential avenues for improvement include:

- **Explicit Rule Induction**: Incorporating mechanisms to induce explicit symbolic rules during training may enhance generalization.
- **Attention Mechanisms**: Leveraging attention-based architectures could help the model focus on relevant parts of the input sequences.
- **Hybrid Models**: Combining neural networks with rule-based engines or employing program synthesis techniques might bridge the gap between neural learning and symbolic reasoning.

Further research is necessary to develop models capable of effective zero-shot reasoning in complex symbolic domains like SPR.