ON THE CHALLENGES OF ZERO-SHOT SYNTHETIC POLYRULE REASONING WITH NEURAL-SYMBOLIC INTEGRATION

Anonymous authors
Paper under double-blind review

ABSTRACT

We investigate the integration of neural networks with symbolic reasoning frameworks to achieve zero-shot learning in Synthetic PolyRule Reasoning (SPR). Despite the potential of neural-symbolic models to generalize to unseen rules without additional training, our experiments reveal significant challenges in generalization and overfitting. Evaluating on the SPR_BENCH benchmark, we observe that traditional neural models struggle to achieve high Shape-Weighted Accuracy (SWA) and Color-Weighted Accuracy (CWA) on unseen rules. These findings highlight the need for more robust neural-symbolic integration methods to realize zero-shot reasoning in SPR.

1 INTRODUCTION

Artificial intelligence has made significant strides in integrating neural networks with symbolic reasoning to tackle complex tasks **?**. Zero-shot learning, the ability of a model to generalize to unseen data without additional training, is crucial for developing adaptable AI systems **?**. Synthetic PolyRule Reasoning (SPR) presents a challenging domain requiring models to infer and apply complex, unseen rules to sequences of symbols.

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 hold promise for generalization, our experiments reveal substantial challenges in enabling models to generalize to unseen rules without further training. We observe that traditional neural approaches tend to overfit to training data and struggle with generalization in the SPR domain.

Our contributions are as follows:

- We implement a neural-symbolic model aimed at zero-shot learning in SPR.
- We evaluate the model on the SPR_BENCH benchmark, analyzing its performance using SWA and CWA metrics.
- We identify key challenges in generalization and overfitting inherent in integrating neural networks with symbolic reasoning for zero-shot SPR.

2 RELATED WORK

Zero-shot learning enables models to classify unseen classes or apply unseen rules without additional training **?**. Neural-symbolic integration combines the learning capabilities of neural networks with the reasoning abilities of symbolic systems **?**. **?** proposed a neural-symbolic system under statistical relational learning, demonstrating potential in zero-shot tasks.

For complex reasoning tasks, **?** introduced a neural-symbolic method leveraging code prompts in large language models. Benchmarks such as KANDY **?** provide datasets for evaluating neuro-symbolic learning and reasoning, similar to SPR_BENCH utilized in our experiments.

3 BACKGROUND

Synthetic PolyRule Reasoning (SPR) involves sequences of symbols governed by underlying rules that determine the correct classification. The SPR_BENCH benchmark provides datasets for training and evaluating models on SPR tasks, focusing on the model's ability to infer and apply rules to sequences.

Metrics: We use Shape-Weighted Accuracy (SWA) and Color-Weighted Accuracy (CWA) to evaluate model performance. SWA weights the accuracy by the variety of shapes in the sequence, while CWA weights by the variety of colors. The PolyRule Harmonic Accuracy (PHA) is the harmonic mean of SWA and CWA, providing an overall performance measure.

4 METHOD

Our approach integrates a neural network with symbolic reasoning to enable zero-shot learning in SPR. The model consists of a neural network component and a symbolic reasoning component. The neural network component is a two-layer Multilayer Perceptron (MLP) that processes sequences of symbols to extract features. Each symbol in a sequence is tokenized into shape and color components, which are encoded into a feature vector.

The symbolic reasoning component uses the extracted features to infer underlying rules and make predictions, aiming to generalize to unseen rules by leveraging the structured representations from the neural network.

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

5 EXPERIMENTS

Our experiments reveal significant challenges in achieving zero-shot learning in SPR through neural-symbolic integration.

Experimental Setup: We conducted experiments using the SPR_BENCH dataset, consisting of training, validation, and test splits. The test set contains sequences governed by rules not seen during training, evaluating zero-shot learning capabilities. The MLP was trained with a maximum of 50 epochs using early stopping based on the PHA metric on the validation set. We used the Adam optimizer with a learning rate of 1×10^{-3}.

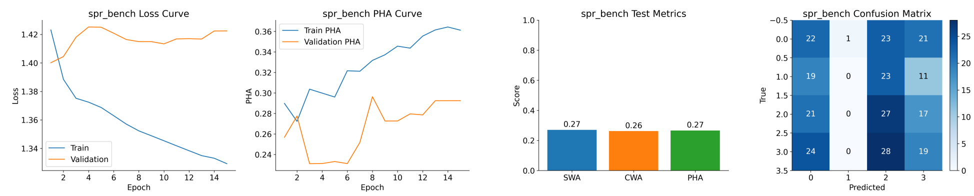
Training Dynamics: As shown in Figure 1 (a), the training loss decreases steadily, indicating learning from the training data. However, the validation loss plateaus and slightly increases after early epochs, suggesting overfitting. The training PHA increases over epochs, but the validation PHA remains low and stable, reinforcing the overfitting concern.

Test Performance: The test metrics are low, with SWA, CWA, and PHA around 0.26–0.27 (Figure 1 (b)). This indicates that the model struggles to generalize to unseen rules in the test set. The confusion matrix reveals significant misclassifications across classes, highlighting that the model has not learned generalized representations applicable to unseen rules.

The results indicate that the neural-symbolic model, as implemented, lacks the ability to infer and apply unseen rules effectively. The significant gap between training and validation performance suggests that the model memorizes training data rather than learning underlying patterns that generalize.

6 CONCLUSION

Our study highlights the 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.



(a) Training Loss and PHA Curves (b) Test Metrics and Confusion Matrix

Figure 1: (a) Training loss decreases steadily, while validation loss plateaus, indicating overfitting. Training PHA increases, but validation PHA remains low. (b) Test metrics (SWA, CWA, PHA) are low, and the confusion matrix shows significant misclassifications, highlighting poor generalization to unseen rules.

Future work should focus on developing more robust neural-symbolic integration methods that can better capture underlying rules in SPR. Techniques such as incorporating explicit rule induction mechanisms, leveraging attention-based architectures, or integrating external symbolic knowledge bases may enhance zero-shot learning capabilities. Exploring alternative neural architectures that can handle sequential data more effectively might also improve generalization in SPR tasks.

References

Supplementary Material

A Additional Experimental Results

A.1 Impact of Joint-Token Representation

We experimented with a joint-token representation where shape and color are combined into a single token. Figure 2 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.

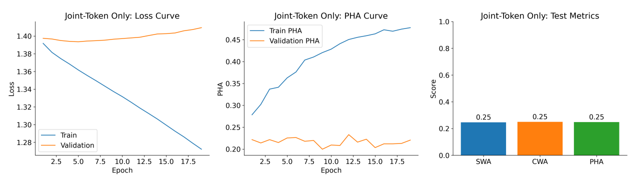


Figure 2: Using a joint-token representation did not enhance generalization; performance remained low with significant overfitting.

A.2 Training Without Early Stopping

We trained the model without early stopping to assess if extended training would aid generalization. Figure 3 shows that the model overfits even more, with training loss continuing to decrease while validation loss increases. The test performance did not improve, indicating that longer training exacerbates overfitting.

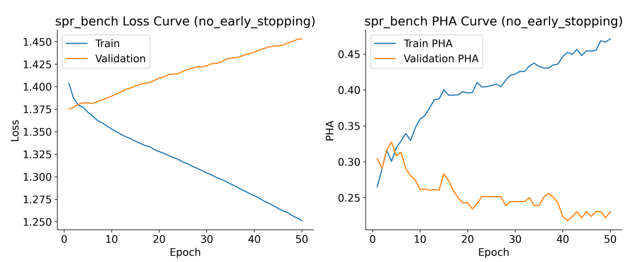


Figure 3: Training without early stopping led to increased overfitting, with validation loss rising while training loss decreased. Test performance did not improve.

A.3 Ablation Studies

Removing Color Features: We trained the model without color features to evaluate the reliance on shape information alone. As shown in Figure 4, performance did not improve, indicating that shape information alone is insufficient for generalization.

Linear-Only Model: We tested a linear model without hidden layers to see if model complexity was contributing to overfitting. Figure 5 demonstrates that the linear model also failed to generalize, suggesting that the issue is not solely due to over-parameterization.

Binary Feature Representation: We experimented with binary feature representations instead of one-hot encoding. Figure 6 shows negligible performance changes, indicating that feature encoding choice is not a significant factor.

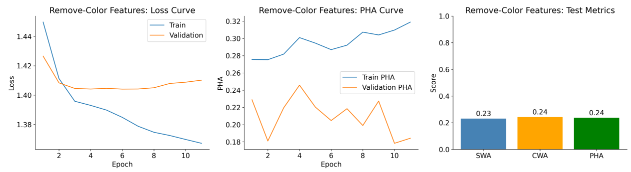


Figure 4: Removing color features did not improve performance, suggesting that both shape and color information are crucial for the task.

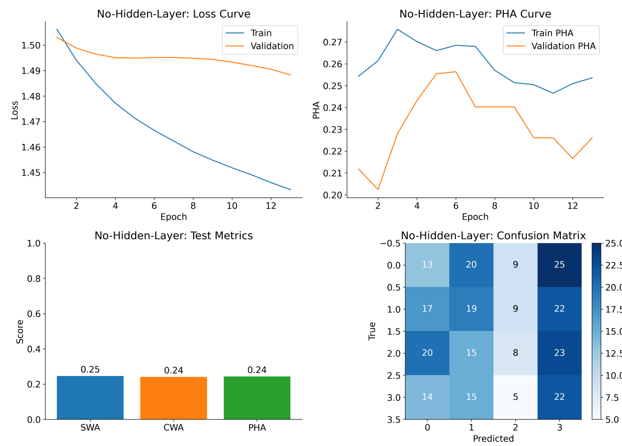


Figure 5: Using a linear-only model did not enhance generalization, indicating that reducing model complexity does not address overfitting.

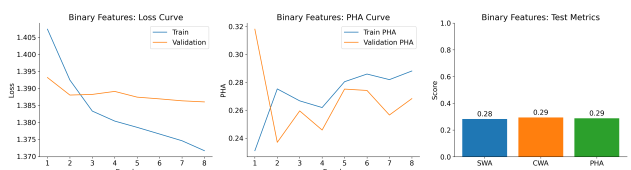


Figure 6: Using binary feature representation did not significantly affect performance, suggesting that feature encoding is not the primary issue.

Length-Invariant Normalization: We applied length-invariant normalization to account for variable-length sequences. As depicted in Figure 7, this adjustment did not yield significant improvements, suggesting that sequence length variability is not the primary issue.

B ADDITIONAL 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 **?**, 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.

One potential avenue for improvement is the incorporation of explicit symbolic reasoning mechanisms that can infer and apply rules more effectively. Additionally, leveraging meta-learning ap-

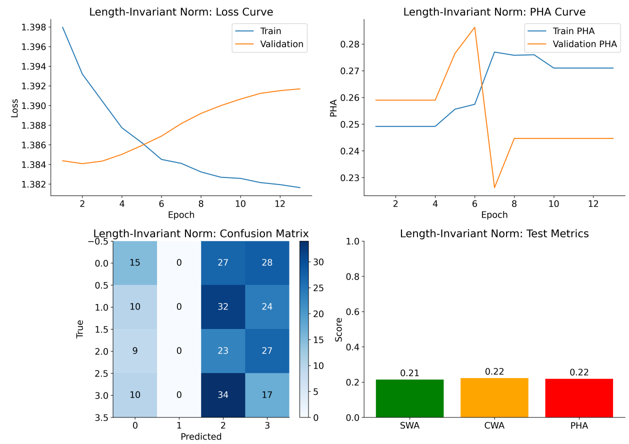


Figure 7: Applying length-invariant normalization showed negligible improvements, indicating that sequence length variability is not the main challenge.

proaches or integrating attention mechanisms might help models focus on relevant parts of the input sequences, enhancing their ability to generalize.

Further research is necessary to develop models that can bridge the gap between neural learning and symbolic reasoning in the context of zero-shot SPR. Exploring hybrid models that combine neural networks with rule-based engines or employing program synthesis techniques may offer promising directions.