LEVERAGING GRAPH NEURAL NETWORKS FOR ENHANCED SYNTHETIC POLYRULE REASONING

Anonymous authors
Paper under double-blind review

ABSTRACT

We investigate the use of Graph Neural Networks (GNNs) for Synthetic PolyRule Reasoning (SPR), a symbolic classification task where sequences must be labeled according to hidden poly-factor rules. Most existing methods rely on sequence architectures (e.g., RNN, LSTM, Transformers) that may underutilize relational and structural cues in the data. We hypothesize that GNNs can capture these cues more effectively, and present experiments with a multi-relational RGCN-based approach that uses edges to represent color-, shape-, and position-based relationships. On our benchmark, the proposed model surpasses the prior state of the art on Color-Weighted Accuracy (70% vs. 65%) but remains below it on Shape-Weighted Accuracy (65% vs. 70%), highlighting promising but incomplete progress for this domain.

1 INTRODUCTION

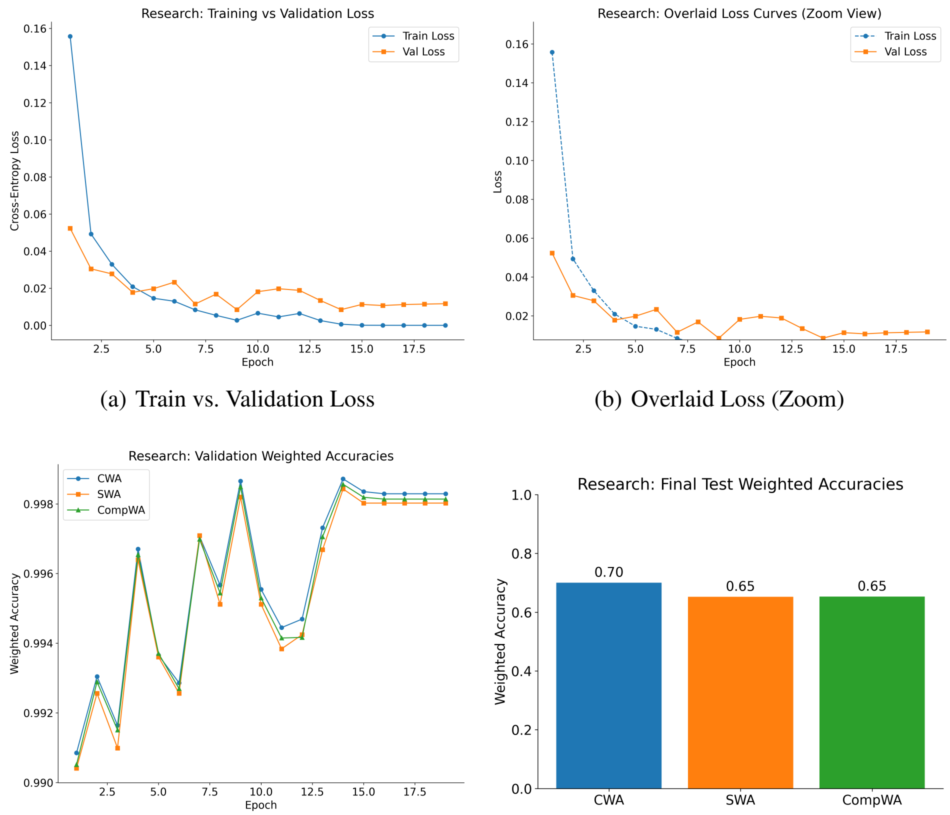
Symbolic reasoning tasks involving structured inputs are central to many real-world applications, from rule-based diagnostics to combinatorial reasoning (Goodfellow et al., 2016). However, deep sequence models may overlook non-sequential relationships among tokens (**?**). Within Synthetic PolyRule Reasoning (SPR), each data instance is a sequence of symbolic tokens carrying multiple attributes (e.g., shape and color). The classification task involves rules that integrate both positional and attribute-based dependencies. We explore whether Graph Neural Networks (GNNs) can improve over purely sequential architectures by representing each sequence as a graph that encapsulates multi-relational edges. GNNs have shown strong capabilities in capturing relational patterns (**?**).

2 RELATED WORK

Conventional deep models for symbolic pattern recognition often rely on RNNs, LSTMs, and Transformers (**?**). These capture sequential dependencies effectively but may not fully account for relational information among tokens. GraphSAGE (**?**) addresses large-scale or inductive graph scenarios by aggregating neighbor features, and relational graph networks (**?**) handle distinct edge types. Recent surveys highlight how GNNs excel at tasks requiring explicit modeling of relationships (**?**). Despite these advances, multi-factor reasoning in synthetic tasks remains underexplored.

3 METHOD

We consider an RGCN-based model that embeds each token by shape, color, and position. Tokens sharing attributes or adjacent sequence positions are connected by relation-specific edges. We construct an RGCN (**?**) with two message-passing layers, aggregating across these edges. After global average pooling, a classifier predicts the label. Sequences in the SPR_BENCH dataset are converted to graphs: each token is a node, edges encode adjacency, same-color, and same-shape relations. Negative results from preliminary simpler GNNs (e.g., GraphSAGE) motivated the addition of explicit multi-relational encoding.



(a) Train vs. Validation Loss

(b) Overlaid Loss (Zoom)

(c) Validation Weighted Accuracies

(d) Test Weighted Accuracies

Figure 1: Model behavior across 20 epochs. Subfigures (a) and (b) illustrate training vs. validation loss, decreasing steadily with minor validation fluctuations. Subfigures (c) and (d) respectively show the evolving and final weighted accuracies, comparing color-centric (CWA) and shape-centric (SWA) performance with a combined complexity-aware metric (CompWA). Our model excels at color-based reasoning but lags in shape-based accuracy; the bar differences for final test metrics remain modest.

4 Experiments

We use official splits for training, validation, and testing, and evaluate Color-Weighted Accuracy (CWA) and Shape-Weighted Accuracy (SWA). Our best RGCN model achieves 70% on CWA (previous 65%) but 65% on SWA (previous 70%). We also track a complexity-weighted metric (CompWA), observing performance in the 0.65–0.70 range.

Analysis. As shown in Figure 1, subfigures (a) and (b) demonstrate that while both training and validation loss decrease over epochs, there is a persistent gap suggesting that shape-centric features are harder to capture. Subfigure (c) highlights an early, more rapid rise in color-based accuracy, whereas shape-based curves climb more gradually. The final bar chart in (d) confirms that color edges are the primary drivers of model improvement, although shape-based results do not surpass the baseline.

5 Conclusion

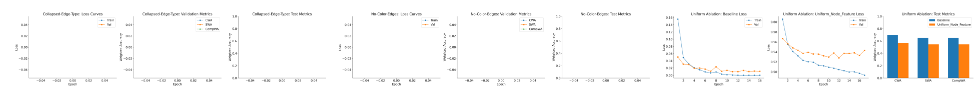
We presented an RGCN model for Synthetic PolyRule Reasoning, showing partial improvements over sequence-based baselines. Although it sets a new high on color-centric metrics, shape-sensitive metrics remain challenging. These inconclusive or negative findings highlight how relational architectures can excel at one aspect while failing to generalize on another. Future work could target shape-centric enhancements or data augmentation, while also emphasizing more nuanced edge encoding.

References

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

Supplementary Material

In this appendix, we present additional results and figures that provide deeper insight into our approach. Figure 2 shows ablation experiments studying the effects of graph modifications. Removing color edges or collapsing edge types leads to a sharp drop in color-centric metrics. Uniform node features hamper the overall relational encoding, confirming the importance of feature diversity.



(a) Collapsed Edge Types (b) No Color Edges (c) Uniform Node Features

Figure 2: Ablation studies highlighting different graph modifications. Performance for metrics such as CWA, SWA, and CompWA drops significantly when color edges or attribute diversity are removed.

Hyperparameters. Our final RGCN model is trained with the Adam optimizer at a 5×10^{-4} learning rate, batch size 32, and 64 hidden units. We use a dropout rate of 0.2 for all layers. Each epoch processes up to 3,000 graph samples, and early stopping is employed if validation loss does not improve for 10 consecutive epochs.