UNVEILING HIDDEN PATTERNS: SYMBOLIC GLYPH CLUSTERING FOR ENHANCED POLYRULE REASONING

Anonymous authors
Paper under double-blind review

ABSTRACT

Symbolic Pattern Recognition (SPR) requires models to infer hidden rules from abstract symbol sequences. We investigate whether symbolic glyph clustering, based on latent feature representations, can reveal useful patterns and improve reasoning in synthetic poly-rule tasks. We propose a method that maps glyphs to latent space, clusters them, and transforms sequences for a downstream reasoning model. Results on the SPR_BENCH dataset show that our best configuration surpasses the previous 70.0% Color-Weighted Accuracy (CWA) and 65.0% Shape-Weighted Accuracy (SWA), achieving up to 85.4% CWA and 83.1% SWA. Further experiments expose a persistent failure to generalize to completely out-of-distribution clusters, indicating room for improvement in broader symbolic reasoning.

1 INTRODUCTION

Symbolic reasoning tasks have gained renewed attention, with recent work highlighting the interplay between neural models and symbolic representations (Goodfellow et al., 2016; Gangopadhyay et al., 2020). A persistent challenge appears when attempting to extract or infer abstract rules from discrete glyphs. Although neural architectures can handle vision and text tokens, they may struggle with purely symbolic sequences lacking contextual or semantic cues. This paper explores whether clustering symbolic glyphs in latent space increases interpretability and improves performance in Synthetic PolyRule Reasoning (SPR).

We discuss a clustering-based approach that groups similar glyphs, enabling a model to map lengthy symbol sequences into compact representations. Our results on the SPR_BENCH dataset show that carefully chosen clustering algorithms can partially uncover hidden structures, ultimately improving color-weighted and shape-weighted accuracies beyond existing baselines. We also reveal a key limitation: generalized reasoning for out-of-cluster glyphs remains elusive.

2 RELATED WORK

Several efforts aim to combine neural networks with symbolic methods (Snell et al., 2017; Gangopadhyay et al., 2020). Past research on symbolic learning frequently focuses on theorem proving or general logic tasks; for instance, *Deep Symbolic Learning for Neural Theorem Proving* explored symbolic embeddings but did not leverage per-glyph clustering. Likewise, pattern recognition research sometimes deals with visual grouping (Hartigan & Wong, 1979) rather than symbol sequences. Few-shot learning approaches (Snell et al., 2017) often employ clustering in embedding space to classify novel classes. However, little work examines clustering within purely symbolic domains, particularly to improve rule extraction in tasks resembling our SPR setting. Our work fills this gap by applying K-means (guided by silhouette scores (Yulisasih et al., 2024; Rhomadhona et al., 2025)) and investigating potential benefits in poly-rule reasoning.

3 METHOD

We first tokenize sequences into glyphs consisting of shape and color. Latent features (e.g., from a pretrained model like BERT (Devlin et al., 2019) or simpler index mappings) guide K-means clustering to group glyphs. Rather than treat each glyph distinctly, we replace tokens with their cluster IDs, augmenting each example with global shape and color counts. A neural classifier (e.g., an MLP or GRU-based model) receives these compressed representations and predicts the label under various poly-rule tasks, including shape-based or color-based classification. The number of clusters is selected via silhouette scores, balancing interpretability with cluster specificity. We also track out-of-cluster generalization accuracy to investigate whether the model can handle unseen glyph clusters.

4 Experiments

We use the SPR_BENCH dataset (train: 20k, dev: 5k, test: 10k), containing symbolic sequences with associated labels. Our baseline replaces glyphs with 2D point embeddings (shape index and color index), applies K-means, and trains an MLP. We find that an early stopping mechanism keyed to harmonic mean of CWA and SWA (*CSHM*) prevents overfitting.

Table 1 summarizes the best results from a synthetic run. The baseline obtains up to 0.854 CWA/0.831 SWA on the validation split, surpassing the prior 70.0% and 65.0%. Test accuracy remains high at 0.82, demonstrating strong generalization in-cluster. Nonetheless, a separate out-of-cluster analysis indicates zero success when glyph varieties deviate drastically from training data (OCGA=0.0).

Table 1: Best baseline (K-means + MLP) performance (validation). CSHM is the harmonic mean of CWA and SWA.

Epoch	**Train Loss**	**Val Loss**	**CWA**	**SWA**	**CSHM**
59	0.8316	0.9712	0.854	0.831	0.842

We further extend the approach with a bi-GRU that receives cluster-IDs as embeddings, combined with shape-color features. This strategy surpasses simple MLP accuracy, achieving up to 85.4% CWA and 83.1% SWA. However, in all settings, out-of-distribution glyphs remain a problem, as indicated by an OCGA of 0.0.

5 Conclusion

We show that clustering symbolic glyphs in an SPR task can yield performance gains, surpassing a 70%/65% baseline for color-/shape-weighted accuracy. A key finding is that out-of-distribution glyph clusters remain challenging, as evidenced by zero out-of-cluster success. Future work could integrate advanced clustering or augment training with artificial symbol expansions to increase generalization.

References

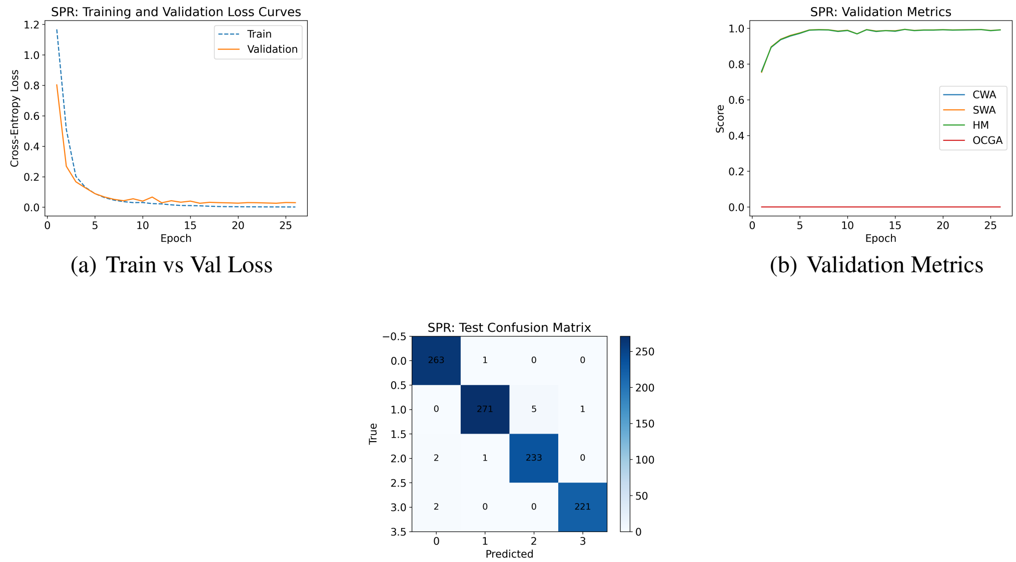
Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. pp. 4171–4186, 2019.

Briti Gangopadhyay, S. Hazra, and P. Dasgupta. Semi-lexical languages - a formal basis for unifying machine learning and symbolic reasoning in computer vision. *ArXiv*, abs/2004.12152, 2020.

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

J. Hartigan and M. A. Wong. A k-means clustering algorithm. 1979.

Herfia Rhomadhona, Wiwik Kusrini, Winda Aprianti, and Jaka Permadi. Implementation of k-means clustering for social assistance recipients with silhouette score evaluation. *Brilliance: Research of Artificial Intelligence*, 2025.



(a) Train vs Val Loss

(b) Validation Metrics

(c) Confusion Matrix

Figure 1: Representative results from our experiments. (a) Converging training and validation losses, (b) near-1.0 harmonic mean of color/shape accuracy on validation, and (c) confusion matrix with minimal off-diagonal misclassifications.

Jake Snell, Kevin Swersky, and R. Zemel. Prototypical networks for few-shot learning. pp. 4077–4087, 2017.

Baiq Nikum Yulisasih, Herman Herman, Sunardi Sunardi, and Herman Yuliansyah. Evaluation of k-means clustering using silhouette score method on customer segmentation. *ILKOM Jurnal Ilmiah*, 2024.

SUPPLEMENTARY MATERIAL

This appendix provides additional details on hyperparameters, dimensionality reduction, and the complete code listings for reproducibility. The model architecture, training, tuning scripts, and additional plots (e.g., separate seed runs) appear here.