Solving Raven's Progressive Matrices via a Neuro-vector-symbolic Architecture

Michael Hersche[1,2], Mustafa Zeqiri[2], Luca Benini[2], Abu Sebastian[1] and Abbas Rahimi[1]

[1]*ETH Zürich, Rämistrasse 101, 8092 Zürich, Switzerland*
[2]*IBM Research-Zurich, Säumerstrasse 4, 8803 Rüschlikon, Switzerland*

Human fluid intelligence is the ability to think and reason abstractly, and make inferences in a novel domain. The Raven's progressive matrices (RPM) test has been a widely-used assessment of fluid intelligence and visual abstract reasoning. Neuro-symbolic AI approaches display both perception and reasoning capabilities, but inherit the limitations of their individual deep learning and symbolic AI components, namely the the so-called neural binding problem and exhaustive symbolic searches explained in the following.

The binding problem in neural networks refers to their inability to recover distinct objects from their joint representation. This inability prevents the neural networks from providing an adequate description of real-world objects or situations that can be represented by hierarchical and nested compositional structures. As an alternative to local representations, distributed representations can provide enough capacity to represent a combinatorially growing number of compositional items; however, they face another issue known as the "superposition catastrophe".
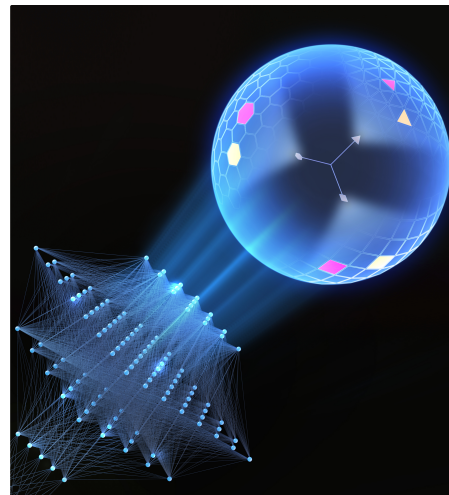

Figure 1: A neuro-vector-symbolic architecture (NVSA): combining neural nets and VSA by leveraging the common language between them namely, the *holographic* vector representations shown as a 3D hologram projection.

The second problem is exhaustive searches in symbolic reasoning. When solving the RPM tests, a probabilistic abduction reasoning approach searches for a solution in a space defined by prior background knowledge which is represented in symbolic form by describing all possible rule realizations that could govern the RPM tests. In effect, the computational complexity of the exhaustive search rapidly increases with the number of objects in the RPM panels, which hinders its utilization for large-scale problems, end-to-end training, and real-time inference.

To address these problems, one viable solution is to exploit vector-symbolic architectures (VSAs), which are computational models where all representations—from atomic to composite structures—are high-dimensional distributed vectors of the same, fixed dimensionality. The

VSA representations can be composed, decomposed, probed, and transformed in various ways using a set of well-defined operations, including binding, unbinding, bundling, permutations, and associative memory.

In [1], we propose a neuro-vector-symbolic architecture (NVSA) consisting of a visual perception frontend and a probabilisitc reasoning backend, both taping into the rich resources of VSA as a general computing framework (see Fig. 1). The resulting NVSA frontend addresses the binding problem in the neural networks, especially the superposition catastrophe, by effectively mapping the raw image of multiple objects to the structural VSA representations that still maintain the perceptual uncertainty. The NVSA backend maps the inferred probability mass functions into another vector space of VSA such that the exhaustive probability computations and searches can be substituted by algebraic operations in that vector space. The VSA operations offer distributivity and computing-in-superposition, which significantly reduce the computational costs thus performing probabilistic abduction and execution in real-time manner.

Results. Compared to the state-of-the-art deep neural network [3] and neuro-symbolic approaches [2], end-to-end training (NVSA e2e tr.) of NVSA achieves a new record of 88.1% in the I-RAVEN dataset (see Table 1). In a fully supervised setting in which the labels of the visual attributes are given (NVSA attr. tr.), the NVSA frontend

Table 1: Average accuracy (%) on the I-RAVEN.

Method	Avg	C	2x2	3x3	L-R	U-D	O-IC	O-IG
PrAE [2]	71.1	83.8	82.9	47.4	94.8	94.8	56.6	37.4
SCL [3]	84.3	99.9	68.9	43.0	98.5	99.1	97.7	82.6
NVSA (e2e tr.)	88.1	99.8	96.2	54.3	100	99.9	99.6	67.1
NVSA (attr. tr.)	99.0	100	99.5	97.1	100	100	100	96.4

can be trained independently of the backend with a novel additive cross-entropy loss, yielding highest accuracy of 99.0%. Moreover, compared to the symbolic reasoning within the state-of-the-art neuro-symbolic approaches, the probabilistic reasoning of NVSA with less expensive operations on the distributed yet transparent representations is two orders of magnitude faster.

Generalization. Further, we analyze the generalization of the NVSA frontend to unseen combinations of attribute values in novel objects. We observe that the frontend with the multiplicative binding cannot generalize to unseen combinations of the attribute values, hence we enhance it by a multiplicative-additive encoding that can generalize up to 72%. The multiplicative binding-based encoding however generalizes well to unseen combinations of multiple objects. We also evaluate the out-of-distribution generalizability of NVSA frontend and backend with respect to unseen attribute–rule pairs. NVSA outperforms the deep learning baselines by a large margin in all unseen attribute–rule pairs.

References

[1] M. Hersche, M. Zeqiri, L. Benini, A. Sebastian, A. Rahimi, A neuro-vector-symbolic architecture for solving Raven's progressive matrices, Nature Machine Intelligence (2023).

[2] C. Zhang, B. Jia, S.-C. Zhu, Y. Zhu, Abstract spatial-temporal reasoning via probabilistic abduction and execution, in: IEEE CVPR, 2021.

[3] Y. Wu, H. Dong, R. Grosse, J. Ba, The scattering compositional learner: Discovering objects, attributes, relationships in analogical reasoning, arXiv preprint arXiv:2007.04212 (2020).