Compositional Generalization in Unsupervised Compositional Representation Learning:

A Study on Disentanglement and Emergent Language

Zhenlin Xu, Marc Niethammer and Colin Raffel
UNC Chapel Hill

THE UNIVERSITY *of* NORTH CAROLINA *at* CHAPEL HILL

Selected as Oral

Compositionality and Generalization

Compositional Generalization: the capability to recognize or generate novel combinations of seen elementary concepts.

✓ Human intelligence

✗ Deep learning system

Related to other generalization problems:

- Domain generalization/OOD (unusual combinations, e.g. a cow on a beach)
- Few-shot/zero-shot (most new samples are new combinations)

Compositionality and Generalization

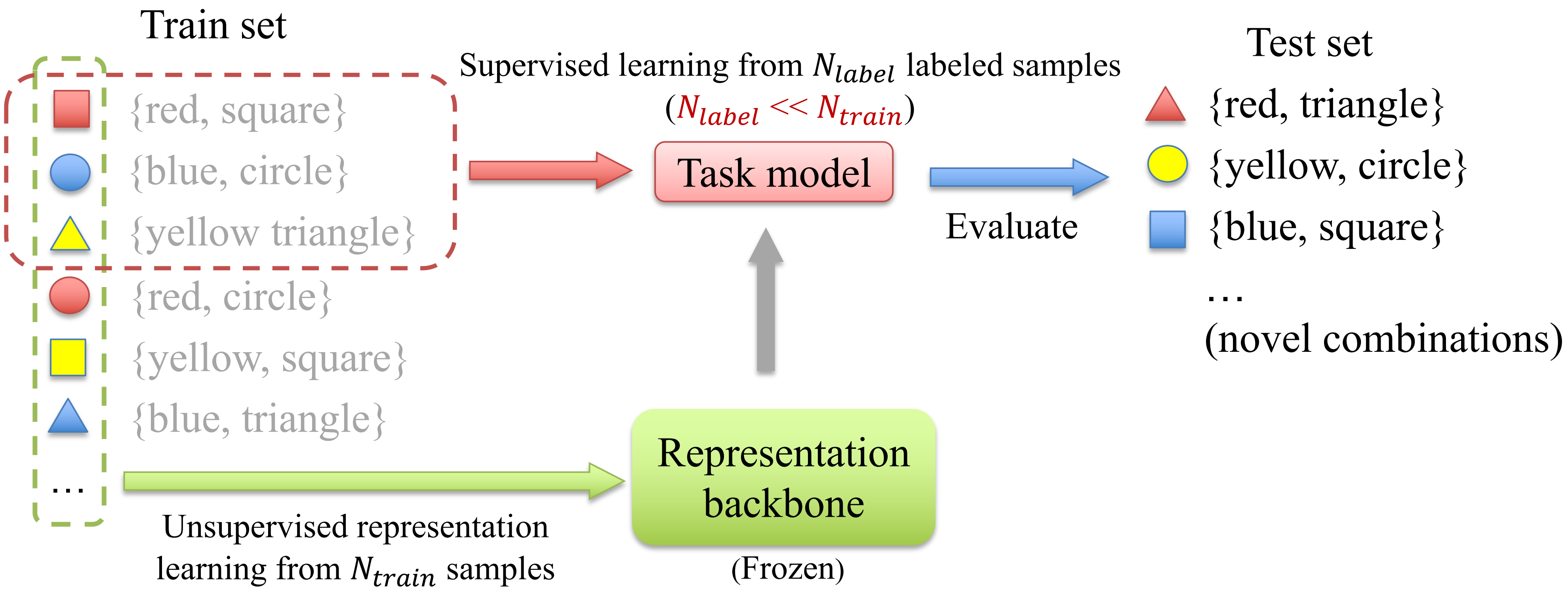
- Human encode complex observations as combinations of primitive representations, aka *Compositional representation.*
  - Represent an object with attributes → color, shape, location …
  - Languages: a sentence → words following grammars

- A common hypothesis: *compositional representations* enable *compositional generalization*.

- We evaluate the above hypothesis on *unsupervised* learning algorithms
  - Disentangled representation learning
  - Emergent language learning

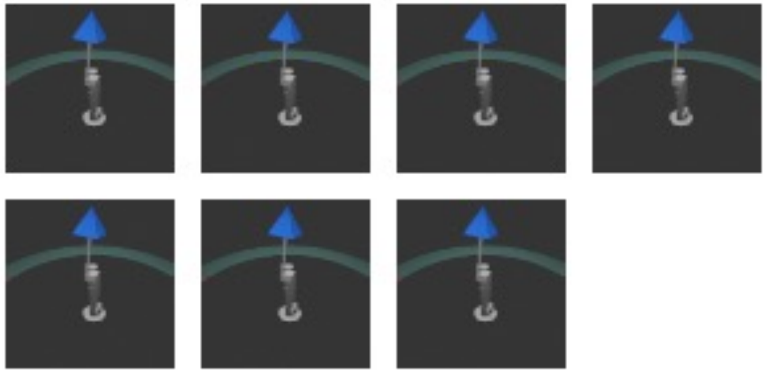
Evaluate Compositional Generalization

- Our criterion: how *easy* to transfer the unsupervised learned representation to downstream tasks with good compositional generalization.
- Evaluation protocol



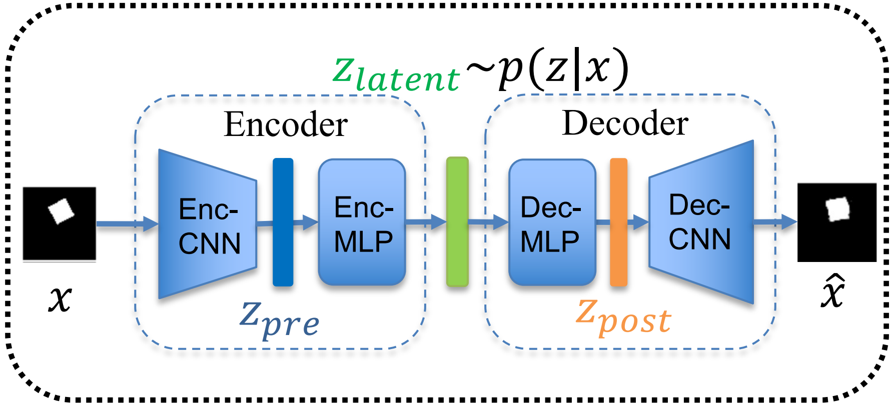
Experimental Setup

- Dataset:
  - dSprites and MPI3D-Real
  - Train/test split: 1 : 9
  - With ground truth categorical and continuous factors
- Unsupervised learning algorithms
  - Disentanglement (β-VAE and β-TCVAE)
  - Emergent language learning
- Downstream tasks:
  - Classification (accuracy) for categorical factors.
  - Regression (R2 score) for continuous factors.
  - Linear and GBT task heads are tested.

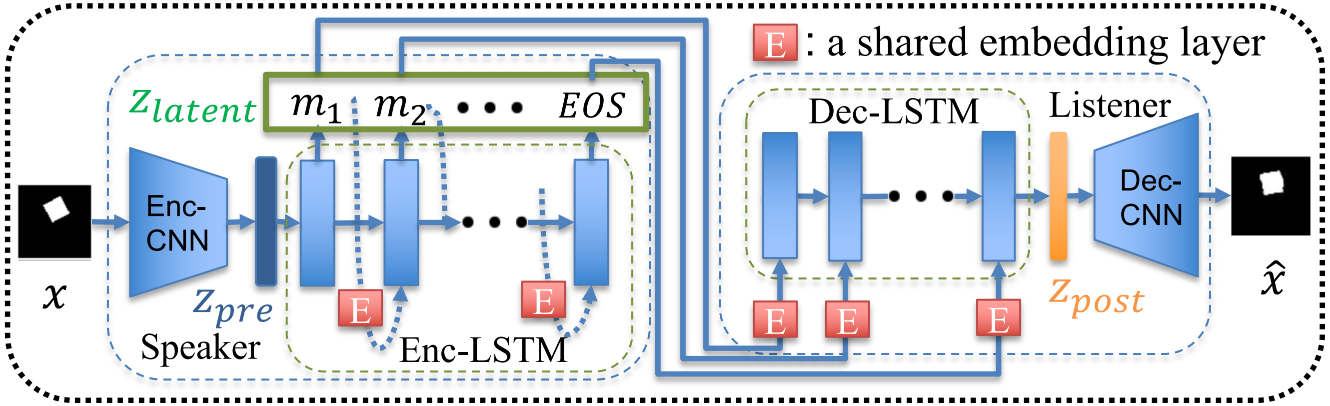


Disentanglement and Emergent Language

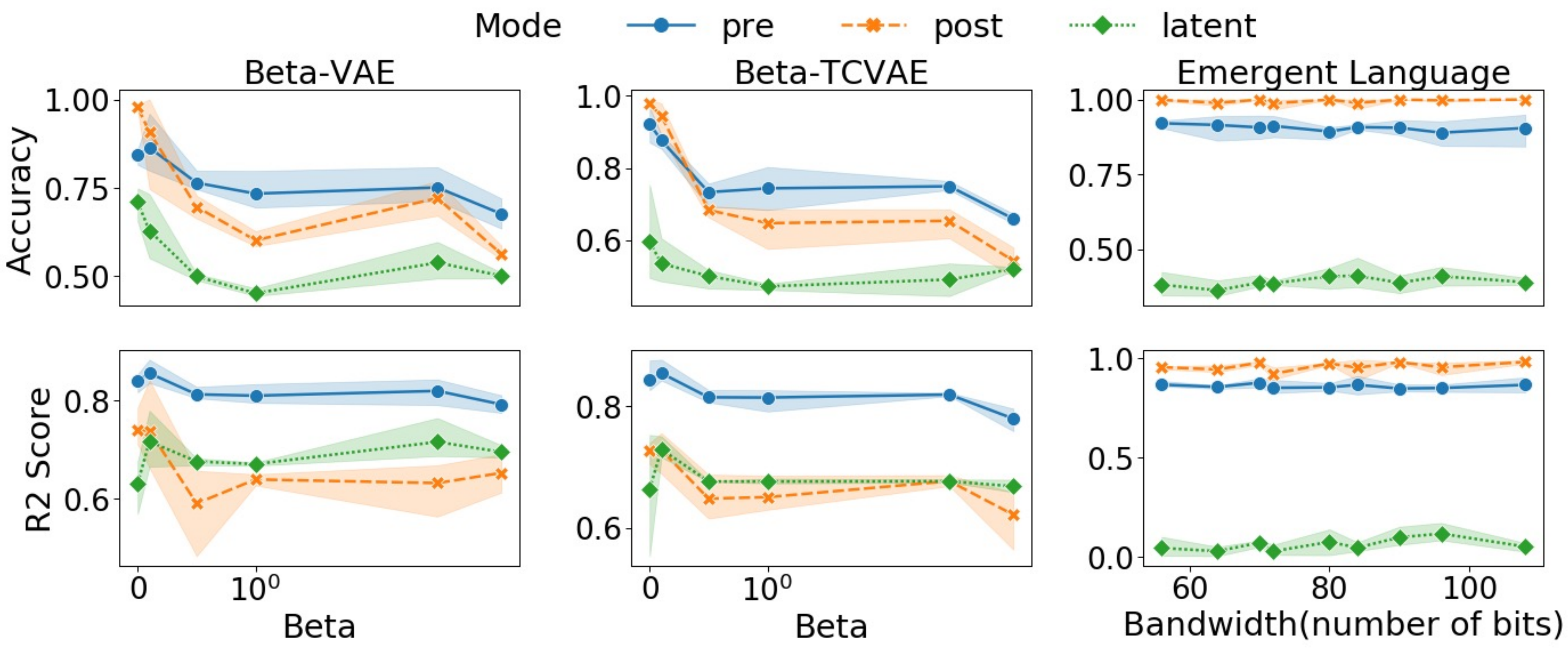
- Disentanglement:
  - "Naive" compositionality
  - Latent variables encode independent factors.
  - Most unsupervised SoTAs are VAE-based.

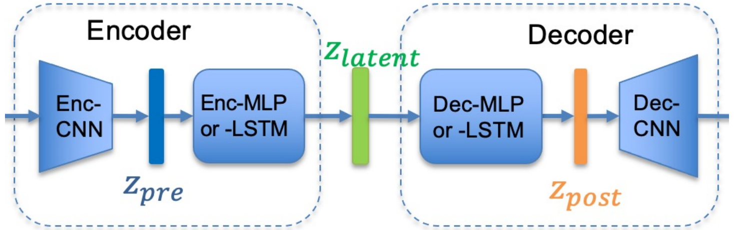


- Emergent language Learning
  - Sequential discrete representations
  - Learning through two-agents communication games.



Finding #1: The Bottleneck Latent Variables are Not Better Representations

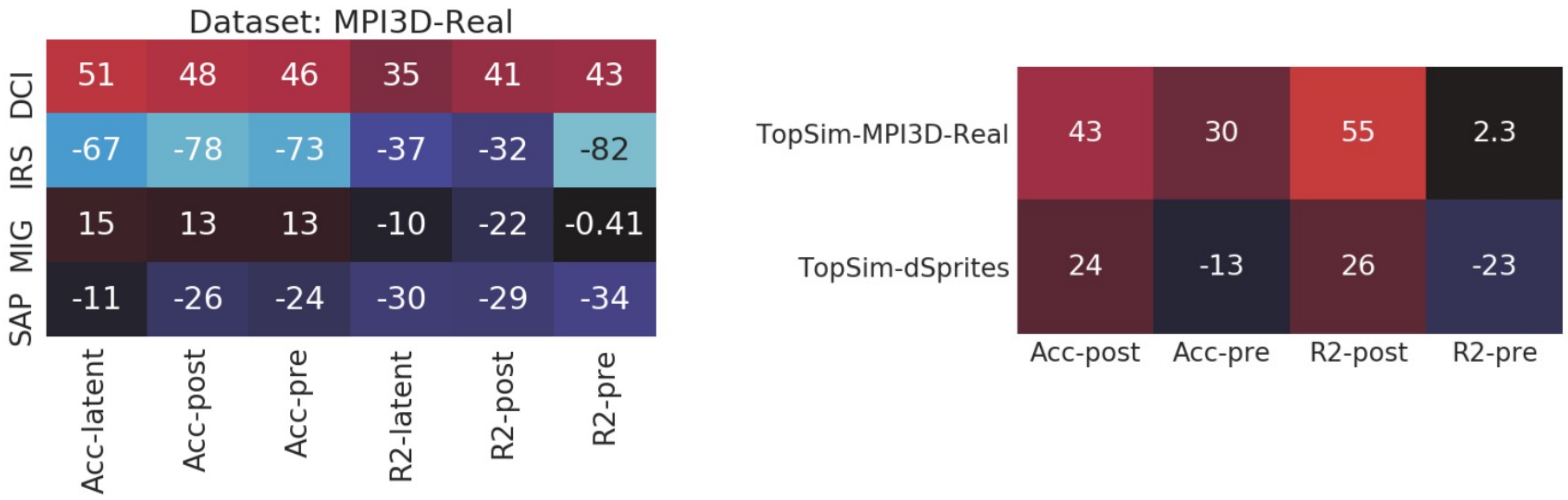




latent << post / pre

Finding #2 Compositionality Metrics != Generalization Performance

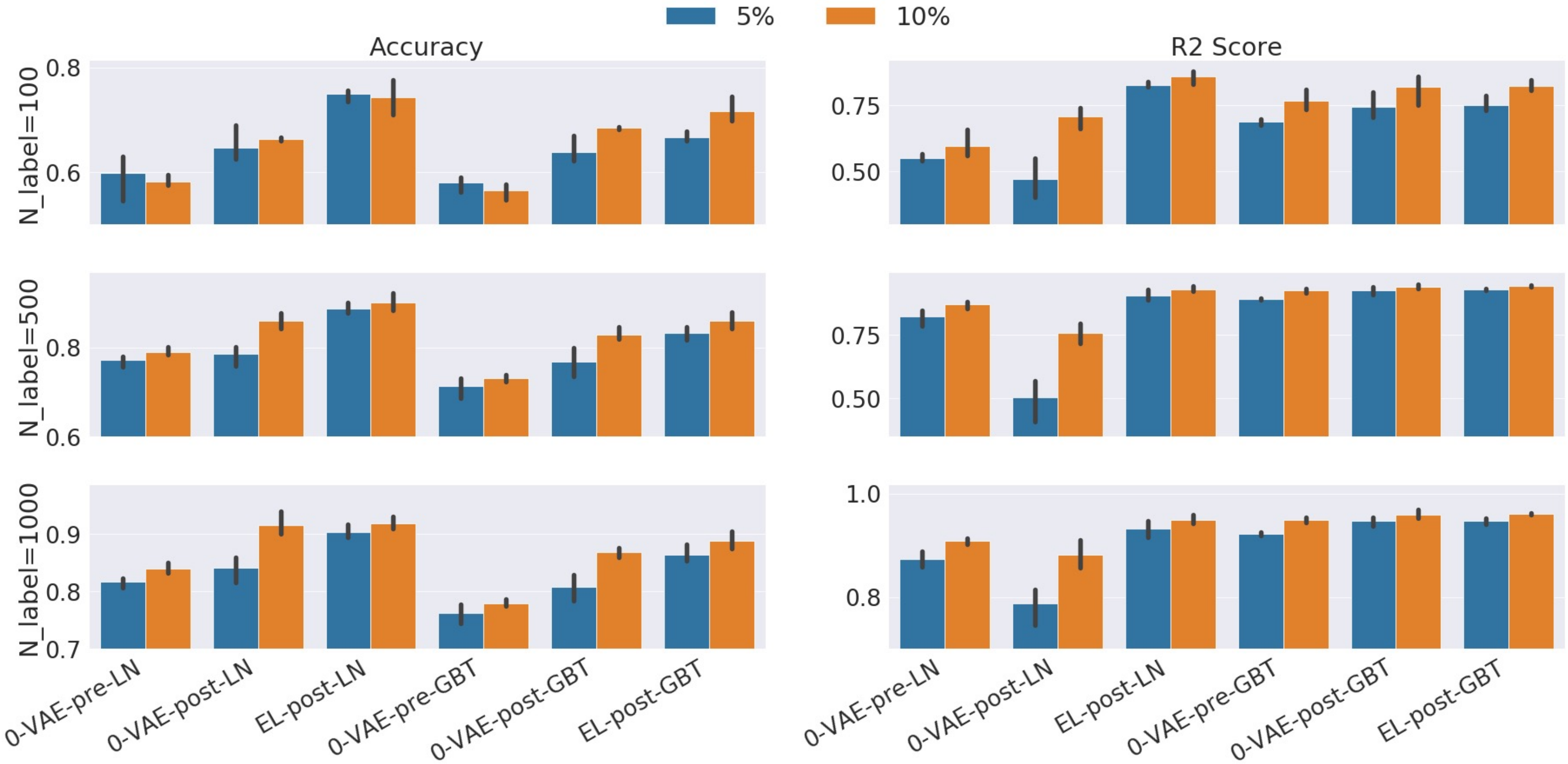
- *None* of compositionality metrics show strong positive correlations with generalization performance.
- Some disentanglement scores even show *negative* correlations.



Ranking correlation between disentanglement scores (left) and topographical similarity (right)

Finding #3 : Emergent Language Generalize Better than Disentangled Representations

- When N_{label} is small, EL-post generalizes significantly better than other models.
- When N_{label} is large, 0-VAE/0-TCVAE may be close to EL-post but classification or regression task favors post/pre and linear/GBT heads differently
- When N_{train} is small (5%), EL-post degrades than 0-VAE/0-TCVAE.



Performances of different representation models under different N_{label} and N_{train} on MPI3D-real

Conclusions

- We proposed a *compositional generalization* evaluation protocol for *unsupervised representation learning* that emphasizes how *easy* we can learn *simple* models for *downstream tasks* with good generalization performance given learned representations.

- Interesting findings:
  - Bottleneck compositional representations do not work well.
  - Compositionality metrics may not imply generalization performance well.
  - Emergent language learning can induce representations with stronger compositional generalization than unsupervised disentanglement learning

Paper: https://arxiv.org/abs/2210.00482

Code: https://github.com/wildphoton/Compositional-Generalization