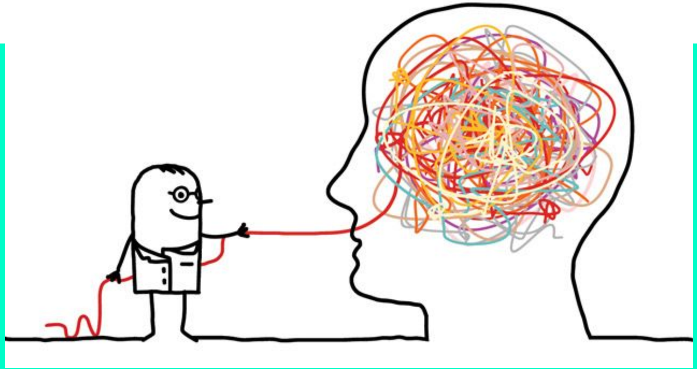
EntaNgled conversations On Disentangled Representations



Organizers

<u>Hosts</u>
Chhavi Yadav : Graduate Student, UCSD @chhaviyadav_
Irina Higgins : Senior Research Scientist, Deepmind

<u>Facilitators</u>
Laure Delisle : Graduate Student, Caltech @laure_delisle
Nivii Kalavakonda : Graduate Student, UW-Seattle @nkalavak

Guests with us today

- Rosemary Nan Ke, PhD Student MILA
- Anirudh Goyal, PhD Student MILA
- Francesco Locatello, PhD Student MPI-ETH
- Niki Kilbertus, PhD Student MPI-Cambridge
- Danilo Rezende, Staff Research Scientist Deepmind
- Stefan Bauer, Research Lead MPI

Outline of the session

- Whiteboarding
- Introduction to the topic
- Discussion with question prompts
- Wrap up

Disentangled Representations (DR)

- Low dimensional
- Capture different sources of variation into disjoint representational subspaces
- Humans & birds do it too!

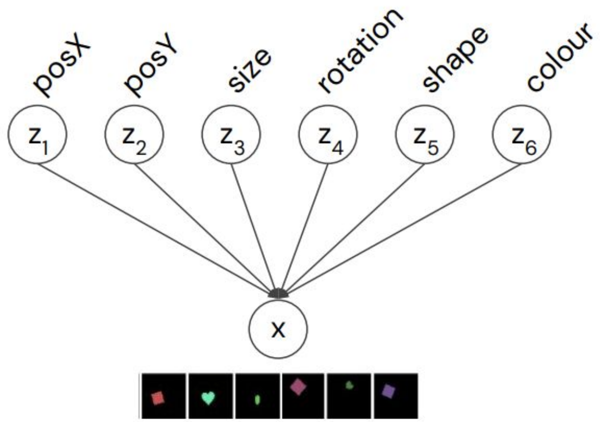


Fig 1. Different factors of variation

Image from here

Video from here

WHY

- Compact & interpretable
- Useful for downstream tasks
- Perform interventions
- Answer counterfactual questions

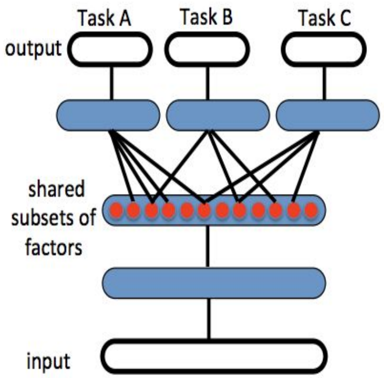


Fig 2. Transfer learning for DR [1]

[1] Bengio et. al. *Representation Learning: A Review and New Perspectives*. IEEE PAMI 2013.

Statistical Perspective

- Product of independent factors
- Colloquial definition, proved impossible w/o supervision & inductive biases on model & data

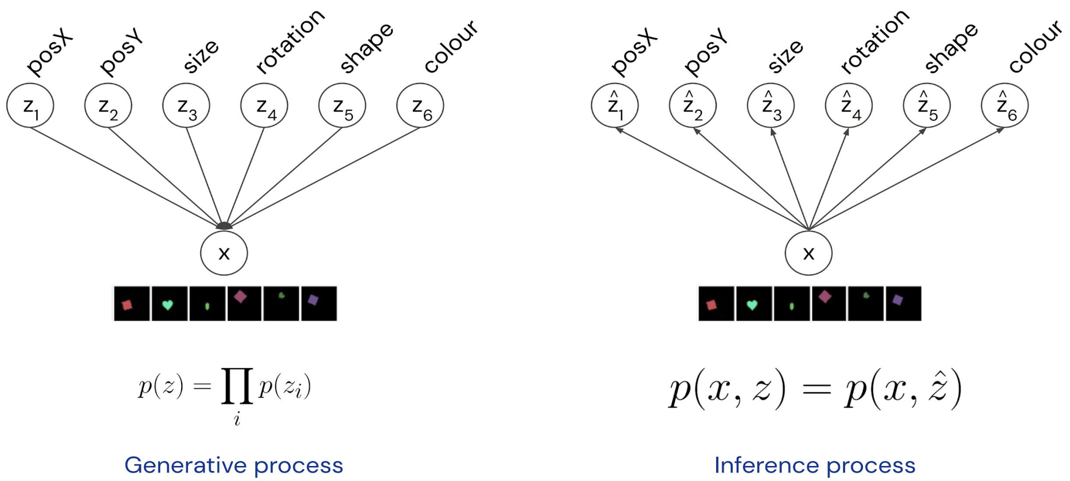


Fig 3. DRs as a product of IFs

Image from here

Causal Perspective

- Learning independent causal mechanisms from data

$$S_i := f_i(\mathbf{PA}_i, U_i), \quad (i = 1, \ldots, n)$$

$$p(S_1, \ldots, S_n) = \prod_{i=1}^{n} p(S_i \mid \mathbf{PA}_i)$$

Independent Causal Mechanisms (ICM) Principle. *The causal generative process of a system's variables is composed of autonomous modules that do not inform or influence each other.*

In the probabilistic case, this means that the conditional distribution of each variable given its causes (i.e., its mechanism) does not inform or influence the other mechanisms.

ICM Principle & equations from [3] Schölkopf et. al. *Causality for Machine Learning.* Arxiv Preprint 1911.10500, 2019.

Symmetry Perspective

- Symmetry group is a set of transformations that leave an object unchanged.
- Square symmetry group
- Closure property

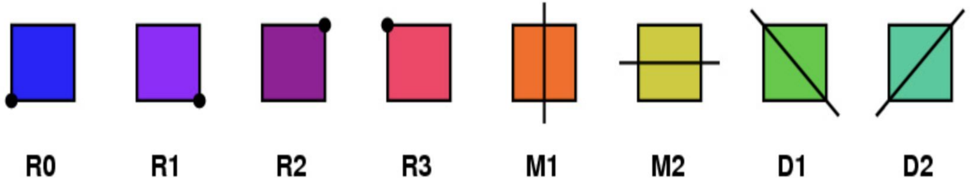
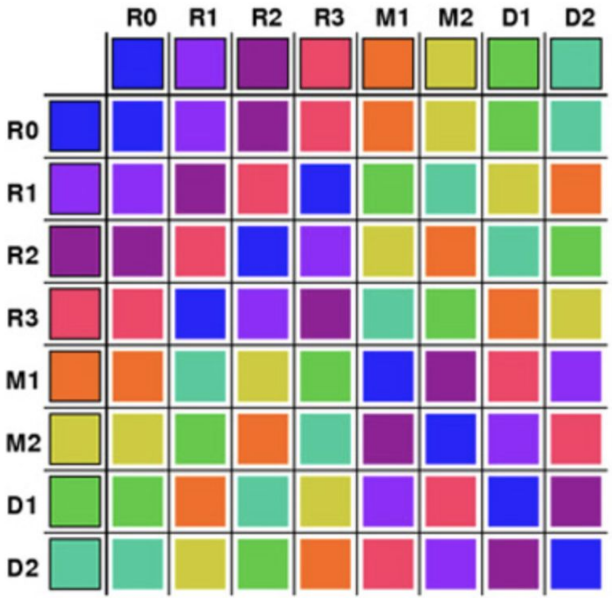


Fig. 4 Symmetries of a Square



Images from [here]

Symmetry Perspective

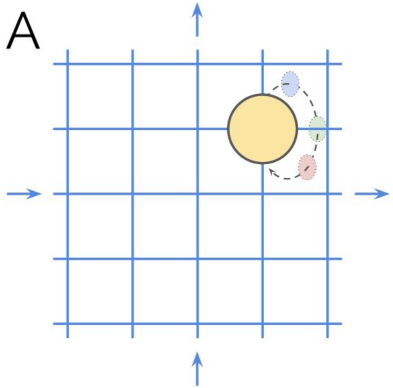


Fig. 5 Gridworld

- Symmetry group --> subgroups
- Each subgroup action affects only one aspect of the world state, keeping others fixed
- $G = G_h \times G_v \times G_c$
- Disentangled representation if it decomposes into independent subspaces, where each subspace is affected by the action of a single subgroup, and the actions of all other subgroups leave the subspace unaffected.*

*[4] Higgins et. al. *Towards a definition of Disentangled Representations*. Arxiv Preprint 1812.02230, 2018.

Food for thought

- Should DRs naturally emerge given enough data?
- What properties should DRs have?
- Is there a single ground truth DR for each dataset?
- How to scale-up the approaches and benchmark them for real-world applications?
- If disentangled subspaces can be multidimensional, do they have to have a particular type of basis?
- We seem to have a “scoring” problem, i.e., many proposed scores don’t capture our intuition of disentanglement well
- How can we deal with survivorship bias when developing unsupervised learning algorithms?
- Do we want to learn a Euclidean manifold or do we need to preserve other kinds of topology?
- Are you doing some interesting stuff during Covid lockdown?
- What are some misconceptions about DRs?

References

[1] Bengio et. al. *Representation Learning: A Review and New Perspectives*. IEEE PAMI 2013.

[2] Locatello et. al. *Challenging common assumptions in the Unsupervised learning of Disentangled Representations*. ICML 2019

[3] Schölkopf et. al. *Causality for Machine Learning.* Arxiv Preprint 1911.10500, 2019.

[4] Higgins et. al. *Towards a definition of Disentangled Representations*. Arxiv Preprint 1812.02230, 2018.

thank you!