

Towards Grounding Conceptual Spaces in Neural Representations

Lucas Bechberger and Kai-Uwe Kühnberger



https://www.lucas-bechberger.de

The Different Layers of Representation



OSNABRÜCK

OSNABRÜCK Dimensions & Domains

- Quality dimensions
 - Different ways stimuli are judged to be similar or different
 - Interpretable by a human
 - E.g., temperature, weight, brightness, pitch
- Domain
 - Set of dimensions that inherently belong together
 - Color: hue, saturation, and brightness
- Distance in this space is inversely related to similarity
- Geometric betweenness represents semantic betweenness
- Concepts: regions in this space

OSNABRÜCK Example: The Color Domain

UNIVERSITÄT



https://en.wikipedia.org/wiki/HSL_and_HSV#/media/File:HSL_color_solid_dblcone_chroma_gray.png

OSNABRÜCK Towards A Thorough Implementation



OSNABRÜCK InfoGAN – Architecture

UNIVERSITÄ

Information Maximizing Generative Adversarial Networks



X. Chen et al., "InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets", Advances in Neural Information Processing Systems, 2016

Towards Grounding Conceptual Spaces in Neural Representations / Lucas Bechberger

OSNABRÜCK InfoGAN – MNIST Results

- Three latent variables
 - Categorical (10 classes)
 - Continuous (uniform)
 - Continuous (uniform)

S .5



X. Chen et al., "InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets", Advances in Neural Information Processing Systems, 2016

Towards Grounding Conceptual Spaces in Neural Representations / Lucas Bechberger

OSNABRÜCK Domains and InfoGAN

Domains in CS framework

- Interpretable dimensions
- Distance-based notion of semantic similarity
- Geometric betweenness represents semantic betweenness

Latent Space of InfoGAN

- Tends to be the case
- Smoothness assumption

 Interpolations in latent space describe a meaningful morph

VERSITÄT OSNABRÜCK Our Proposal

- Use InfoGAN to learn dimensions of "difficult" domains
 - For starters: shape domain
- Hyperparameters
 - Number of latent variables
 - \rightarrow as few as possible, as much as necessary
 - Type of latent variables
 - \rightarrow continuous (i.e., uniform or Gaussian)
- Make sure that we learn **new** dimensions
 - Pre-select training data (if possible)
 - Additional term in loss function: correlation to other dimensions

First Preliminary Results

- Data set of right-angled triangles, rectangles, and ellipses
- 2 continuous varibles (uniform distribution), 500 epochs



OSNABRÜCK Towards A Thorough Implementation



Overall Envisioned System

OSNABRÜCK

UNIVERSITÄT





Thank you for your attention!

Questions? Comments? Discussions?



https://www.lucas-bechberger.de