

---

# Unsupervised Learning of Disentangled and Interpretable Representations from Sequential Data

## Report

---

**Stefan Wezel**

stefan.wezel@student.uni-tuebingen.de

4080589

ML4S

## Abstract

Information in sequential data is often distributed over multiple time scales. While if viewed as a single signal, such data might appear noisy. However, patterns can emerge if temporal scales are viewed separately from one another. Hsu et al. [1] leverage this intrinsic structure to learn disentangled representations from sequential data in an unsupervised manner. They propose to factorize sequence level and segment level attributes into distinct latent subspaces. Architectural and sequence dependent priors create an inductive bias to encourage the proposed factorization. Here, we put their work into a formal context, explore the proposed methodology, and reflect critically on their work.

## Introduction

Intuitively, disentangled representations are reflective of the underlying generating factors of observed data in thus they are encoded as separate latent subspaces. This notion is already present in classical factor analysis work, where it is referred to as independent component analysis (ICA) [2]. However, many problems cannot be solved in linear fashion. The vast success of deep neural networks (DNN) can be largely attributed to the fact that they are very powerful non-linear function approximators. Thus, making them an promising method to solve the long standing problem of non-linear ICA.

Different methods have been proposed to learn such disentangled representations [3, 4, 5] with varying success. Many of these works focus on image data. However, it has been shown by Locatello et al. [6] that disentangled representations cannot be learned without introducing any kind of supervision or inductive biases. Sequential data, while having been explored less, despite offers inherent structure that can be exploited to construct inductive biases as has been proposed by Hsu et al. [1]. Besides technical challenges, this strain of research suffers from the lack of formally defined and agreed upon foundations. The very term of disentangled representations for example is often understood differently in between works. In the following section, we will use the definition, proposed by Higgins et al. [7] to put the work by Hsu et al. [1] into formal context.

## Viewing the FHVAE though a Formal Lense

Works on disentanglement often lack a proper formal framework [7]. While claiming to achieve disentanglement, Hsu et al. [1] do not provide any formal foundation for their claim. With no formal tools at hand, we cannot formally discuss whether disentanglement was achieved. Thus, we use the following sections to respectively give a theoretical background that defines disentanglement in group theoretic terms, and describe the problem, the FHVAE was designed to solve, in these terms.

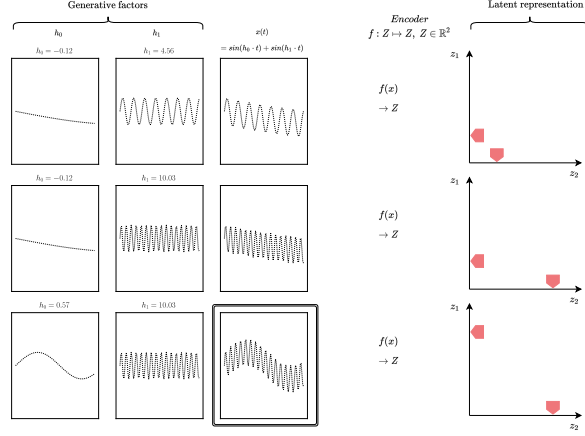


Figure 1: Changing values of generating factors are reflected in the corresponding latent variables.

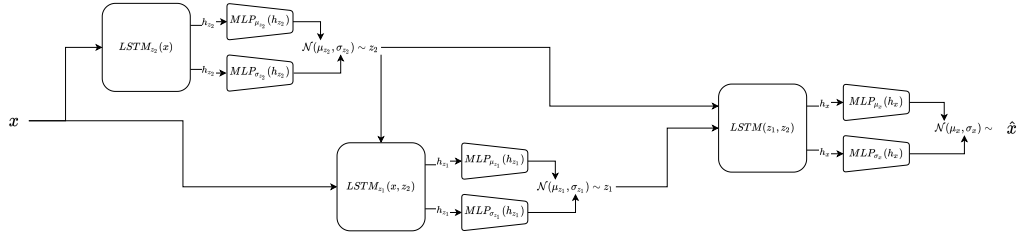


Figure 2: Architecture of the proposed FHVAE.

## The Tools of Group Theory

A group is described by a tuple of an operation  $\circ$  and a non-empty set  $G$ . The set has to be closed under the operation, it must contain an identity element, the operation must be associative, and for every element in  $G$ , there must be an inverse element [8]. Groups can be composed via a direct product. The result of a direct product of groups is itself a group.

A group that is of particular interest for the field of disentangled representation learning is the symmetry group [8]. A symmetry group consists of a set of transformations that leave a given object  $X$  invariant and the operation is the composition of such transformations. A prominent example of a symmetry group is  $SE_3$ . This symmetry group can be visualized as a set set of vertices  $X$  that form an equilateral triangle. Permutations of this set would result in rotations or flips of the triangle. These permutations would be the symmetry transformations of our symmetry group. The operation would be composing multiple permutations.

Another important concept is the group action. It is the result of applying a symmetry transformation to an object. In our triangle example, a group action would be the permuted set of vertices.

## Symmetries in Sequential Data

### FHVAE - Constructing an Equivariant Map

#### Results

#### Discussion

Lack of formal context. Lack of evidence for disentanglement. Further exploit the available data using cross reconstruction.

#### Conclusion

## References

- [1] Wei-Ning Hsu, Yu Zhang, and James Glass. Unsupervised learning of disentangled and interpretable representations from sequential data. *arXiv preprint arXiv:1709.07902*, 2017.
- [2] Pierre Comon. Independent component analysis, 1992.
- [3] Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. beta-vae: Learning basic visual concepts with a constrained variational framework. 2016.
- [4] Xi Chen, Yan Duan, Rein Houthoofd, John Schulman, Ilya Sutskever, and Pieter Abbeel. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. *arXiv preprint arXiv:1606.03657*, 2016.
- [5] Tejas D Kulkarni, Will Whitney, Pushmeet Kohli, and Joshua B Tenenbaum. Deep convolutional inverse graphics network. *arXiv preprint arXiv:1503.03167*, 2015.
- [6] Francesco Locatello, Stefan Bauer, Mario Lucic, Gunnar Raetsch, Sylvain Gelly, Bernhard Schölkopf, and Olivier Bachem. Challenging common assumptions in the unsupervised learning of disentangled representations. In *international conference on machine learning*, pages 4114–4124. PMLR, 2019.
- [7] Irina Higgins, David Amos, David Pfau, Sebastien Racaniere, Loic Matthey, Danilo Rezende, and Alexander Lerchner. Towards a definition of disentangled representations. *arXiv preprint arXiv:1812.02230*, 2018.
- [8] Georgia Benkart. Abstract algebra, by in herstein. *The American Mathematical Monthly*, 94(8): 804–806, 1987.