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 by 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.

1

Viewing the FHVAE though a Formal Lense

The Tools of Group Theory

Symmetries in Sequential Data

FHVAE - Constructing an Equivariant Map

Results

Discussion

Lack of formal context. Lack of evidence for disentaglement. 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 Houthooft, 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.