Do top-down predictions of time series lead to sparse disentanglement?

Kosuke Miyoshi (1,2,3), Naoya Arakawa (2,3), Hiroshi Yamakawa (2,3,4)
Narrative Nights Inc.(1), The Whole Brain Architecture Initiative(2),
Dwango Artificial Intelligence Laboratory(3), The University of Tokyo(4)

Abstract

A framework that defines functions and interface semantics of the cortical micro-circuit was previously proposed as The Coritical Master Algorithm Framework (Yamakawa 2017), and a deep learning model that supports this framework is demanded.
We chose the VRNN network model (Variational Recurrent Neural Network) to add a hierarchical feature. We implemented time series prediction with top-down signal, and found the representation in the lower layers became sparsely disentangled, so that the fundamental factors in the sensor input were extracted. We also discuss the ability of functional differentiation with HVRNN.

Cortical Master Algorithm Framework

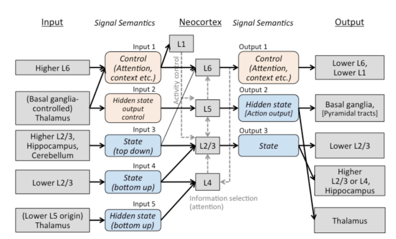
A framework that defines the functions and interface semantics of cortical micro-circuits was previously proposed as the Cortical Master Algorithm Framework (MAF).

MAF requires functional and structural prerequisites as below.

Functional prerequisites

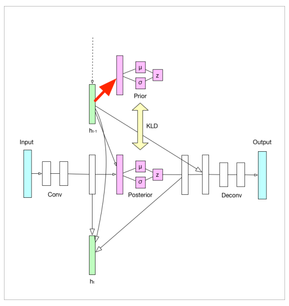
- Dimension reduction
- Unsupervised learning
- Time series prediction
- Disentanglement
- Generative model
- Sparsity
- Internal state
- Hierarchy
- Prediction with top-down signal
- Attention with control signal

Structural prerequisites

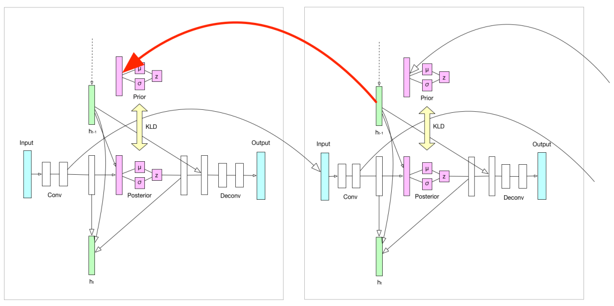


Hierarchical VRNN (HVRNN)

VRNN



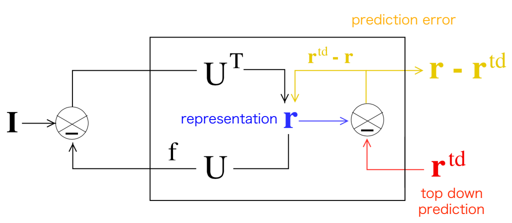
HVRNN (Proposed)



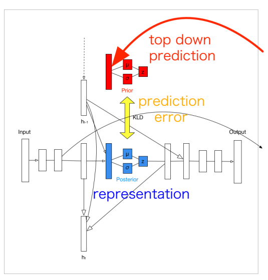
VRNN (Variational Recurrent Neural Network) (Chung 2015) is a deep learning model that is capable of dimension reduction, unsupervised learning, time series prediction, generative model, and internal state. VRNN has the structure in which time series prediction is processed with latent variable **z**. Latent variable **z**'s approximated posterior distribution $\mathbf{q(z_t|x_{\le t},z_{<t})}$ is predicted by the prior distribution $\mathbf{p(z_t|x_{<t},z_{<t})}$.

By following the MAF requirements, we propose a Hierarchical VRNN model (HVRNN) that adds hierarchy to the VRNN. In HVRNN we implemented time series prediction in lower layers with top-down signals by following predictive coding (Rao 1999).

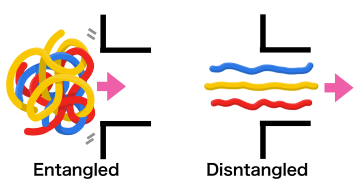
Predictive Coding (Rao & Ballad, 1999)

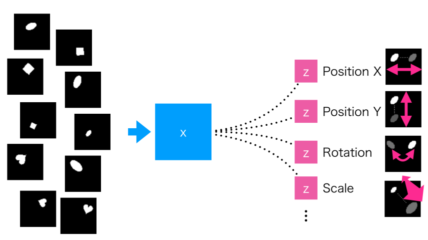


HVRNN



Disentanglement



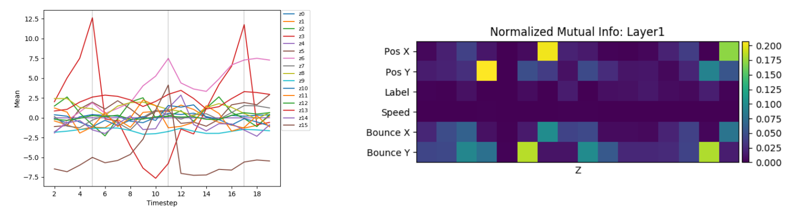


One of the MAF functional prerequisites is disentanglement. Higgins et al. treat the function of redundancy reduction and extraction of the statistically independent parameters that are supposed to exists in the visual ventral pathway as the disentanglement of factors (Higgins 2016).

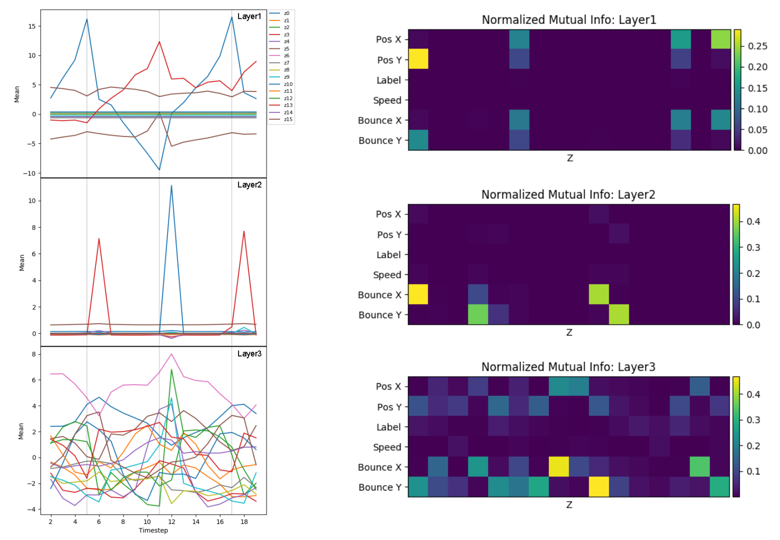
Experiments

To examine HVRNN, we trained both VRNN and HVRNN using our own dataset following the Moving MNIST format with unsupervised training. One sequence consists of a 20 step image time series (64x64 pixels).

VRNN



HVRNN

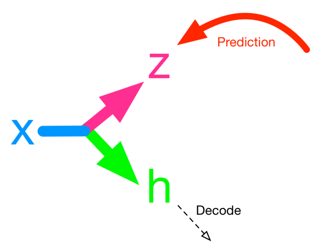


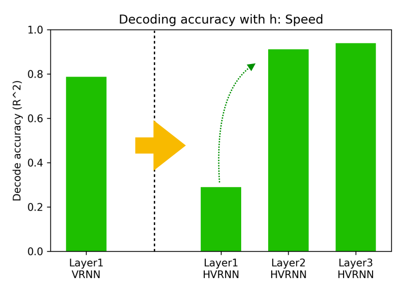
The graphs on the left show the mean of latent variable **z**'s posterior during 20 time steps. Figures on the right show the normalized mutual information between each dimension of **z** and the factors that the input image data comprises.

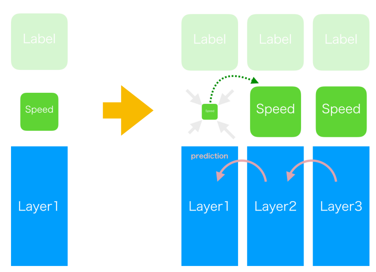
In VRNN some dimensions of latent variable correspond to the movement of a digit, and in HVRNN this correspondence becomes more apparent, and they are more sparsely disentangled in the first layer. In HVRNN we can clearly see that in the second layer **z** corresponds to the timing of the bounces of the digit at the top and bottom walls with precise one frame delay after the bounce timing.

Discussion: Functional Differentiation

VRNN has the structure like the conditional VAE which is conditioned on RNN hidden state **h**. Conditional VAE tries to find factors of input **x** that are not included in the condition **h** and extract them as latent **z**. In addition to the disentanglement analysis of **z**, we also investigated the factor decoding ability of RNN hidden state **h**







The graph on the left shows the decoding accuracy (R^2 score) of the speed factor trained by Ridge regression with RNN state **h**. Single layer VRNN has the ability to decode speed factor at a high rate, but when the top-down prediction was added it decreased and upper layers acquired higher accuracy. Decoding accuracy of the label factor didn't change. This result implies the possibility of functional differentiation caused by the top-down prediction in HVRNN structure.