

Fast Sparse Coding for Spatio-temporal Prediction

Jonathan S. Yedidia
Disney Research Boston,
222 Third St., Cambridge, MA 02142 USA
yedidia@disneyresearch.com
<http://jonathanyedidia.wordpress.com>

A functioning module in a biological or artificial brain must process an on-going stream of spatio-temporal inputs, and learn to predict those inputs. I formalize the problem of learning to make accurate predictions of future space-time inputs obtained from an unknown “world,” and then describe a system that can quickly solve this problem, at least for relatively simple synthetic worlds.

The system is based on modules implementing sparse coding, where each module is endowed with a short-term memory which gives it access to a discretized space-time image consisting of spatial images for the present and a small number of previous “frames.” A module attempts to recreate an input space-time image as a weighted sum of a small number of basis vectors. For efficiency, all space-time images, weights, and basis vectors are required to be strictly non-negative. All processing is strictly on-line; each space-time image is used, learned from, and discarded. A module has no knowledge about the number of basis vectors that should be used to recreate a particular space-time input, or the number required to describe the world.

A module first infers weights describing the contributions of existing basis vectors to a space-time image, using a matching pursuit algorithm. It then learns and adapts its set of basis vectors. The active basis vectors for a space-time image are projected part way towards the values they should have to reproduce the image. Additionally, basis vectors are recruited to reproduce missing parts of the image, and deleted if they are inactive or redundant. Basis vectors are merged if they are “time-sharing,” and small basis vector elements are clipped to zero.

In this framework, prediction is easy. Once an accurate set of basis vectors has been learned, one can predict the future by shifting one frame into the future in a space-time image, inferring the basis vectors as usual using matching pursuit, and filling in the unknown next frame. The process can be iterated to predict many frames into the future. Prediction thus becomes erasure-correction of the missing future.

I will show software demonstrating that such a module can very quickly learn

and adapt to its environment, and predict its inputs, and might thus serve as a building block module for an artificial brain. The surprising speed of the sparse coding algorithms is a result of the non-negativity requirements on the images, weights, and basis vectors, and the use of recruitment. In a sequential version of the software, the time complexity of the inference and learning algorithms scales linearly with the number of basis vectors in a module, but the algorithms used are embarrassingly parallel, and if parallelized, would take time that would grow only very slowly with the number of basis vectors.

I will describe how these modules can be composed into networks, and speculate on how control algorithms might achieve goals using a related erasure-correction algorithm. Finally, I will relate the various ingredients in the system to their counterparts in biological brains.

Topics: learning algorithms; prediction and sequence modeling

Preference: oral