Learning large-scale heteroassociative memories in spiking neurons

Aaron Russell Voelker (arvoelke@uwaterloo.ca)
Eric Crawford (e2crawfo@uwaterloo.ca)
Chris Eliasmith (celiasmith@uwaterloo.ca)
Centre for Theoretical Neuroscience, University of Waterloo, Waterloo, ON, N2L 3G1

Abstract

Associative memories have been an active area of research over the last forty years (Willshaw et al., 1969; Kohonen, 1972; Hopfield, 1982) because they form a central component of many cognitive architectures (Pollack, 1988; Anderson & Lebiere, 1998). We focus specifically on associative memories that store associations between arbitrary pairs of neural states. When a noisy version of an input state vector is presented to the network, it must output a "clean" version of the associated state vector. We describe a method for building large-scale networks for online learning of associations using spiking neurons, which works by exploiting the techniques of the Neural Engineering Framework (Eliasmith & Anderson, 2003). This framework has previously been used by Stewart et al. (2011) to create memories that possess a number of desirable properties including high accuracy, a fast, feedforward recall process, and efficient scaling, requiring a number of neurons linear in the number of stored associations. These memories have played a central role in several recent neural cognitive models including Spaun, the world's largest functional brain model (Eliasmith et al., 2012), as well as a proposal for human-scale, biologically plausible knowledge representation (Crawford et al., 2013). However, these memories are constructed using an offline optimization method that is not biologically plausible. Here we demonstrate how a similar set of connection weights can be arrived at through a biologically plausible, online learning process featuring a novel synaptic learning rule inspired in part by the well-known Oja learning rule (Oja, 1989). We present the details of our method and report the results of simulations exploring the storage capacity of these networks. We show that our technique scales up to large numbers of associations, and that recall performance degrades gracefully as the theoretical capacity is exceeded. This work has been implemented in the Nengo simulation package (http://nengo.ca), which will allow straightforward implementations of spiking neural networks on neuromorphic hardware. The result of our work is a fast, adaptive, scalable associative memory composed of spiking neurons which we expect to be a valuable addition to large systems peforming online neural computation.

Our network consists of three layers of spiking neurons. The first layer encodes the current input, and the third layer encodes the current output. Associations are learned by changing the efferent and afferent connection weights of the middle layer using biologically plausible learning rules.

1 Input Learning Rule

Given a learning rate η, an input vector x encoded by the activity of the input layer, the filtered activity $a(t)$ of neurons in the middle layer, and the matrix $\mathbf{e}$ whose rows are the "preferred direction" vectors of the middle layer neurons, we modify the preferred direction vectors of the middle layer neurons according to the equation:

$$\Delta\mathbf{e} = \eta(a(t)x^T - a(t)\mathbf{e}) = \eta a(t)(\begin{pmatrix} x^T \\ \vdots \\ x^T \end{pmatrix} - \mathbf{e}) \tag{1}$$

Setting $\Delta\mathbf{e}_i = 0$ gives $a_i(t)x = a_i(t)\mathbf{e}_i$, and thus for a particular x, stability is characterized by

$$[\Delta\mathbf{e}_i = 0] \iff [a_i(t) > 0 \implies \mathbf{e}_i = x] \tag{2}$$

Thus the effect of this rule is to make a subset of the middle layer neurons fire only when x or a noisy version of it, is presented. The preferred direction vectors of the middle layer neurons are encoded in the synaptic weights between the first and middle layer neurons. Consequently, changing the preferred direction vectors corresponds to changing the connection weights using a local learning rule which we will outline in detail in the finished version.

2 Figures

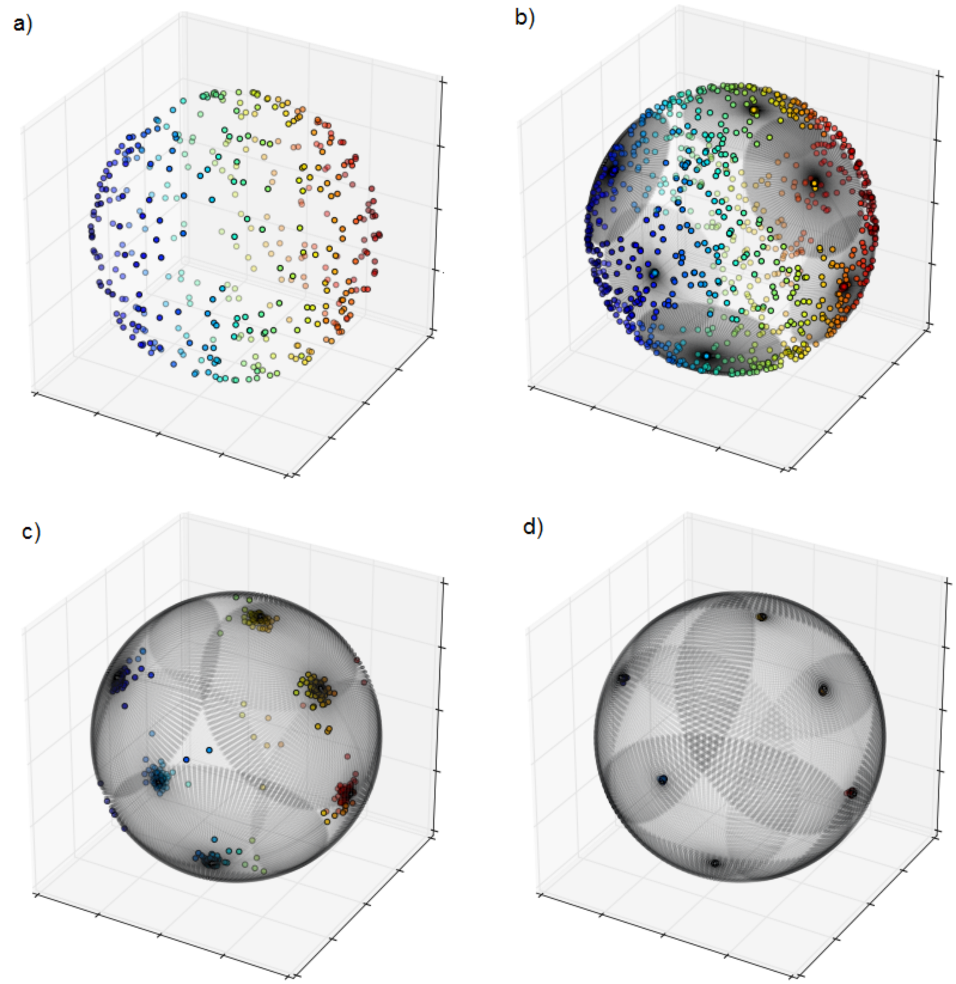

Figure 1: Input vector clustering of a 3-dimensional ensemble, presented with 6 evenly spaced x. (a) The initial clustering, before any input has been given. (b) Each input vector has been shown for 2 seconds of simulation time, with intercepts set to $cos(\frac{\pi}{6})$. Gray plates show the area of attraction for each input vector. (c) Same as b, with intercepts chosen to give $p_i = \frac{1}{6}$. (d) Same as b, with intercepts set to $cos(\frac{\pi}{3})$.

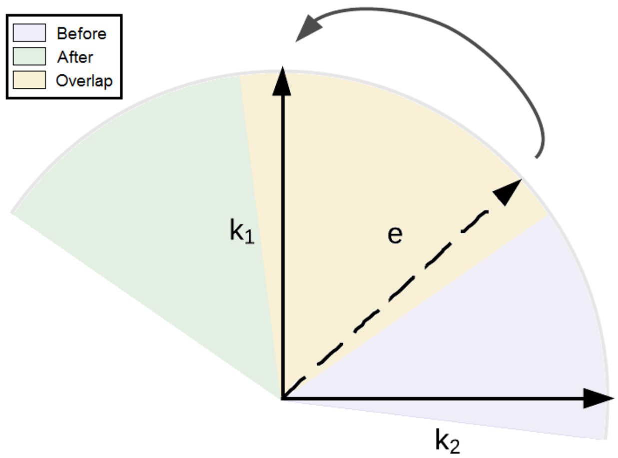

Figure 2: A two-dimensional 'preferred direction' (or 'encoding') vector of a neuron, e, is shifted to the first input vector to be presented, k_1, where it will not respond to k_2. This required that $k_1 \cdot k_2 < c$, where c is the intercept of e. The intercept of the neuron is depicted by shading the responding region. Note that before e was shifted, it was sensitive to both k_1 and k_2.

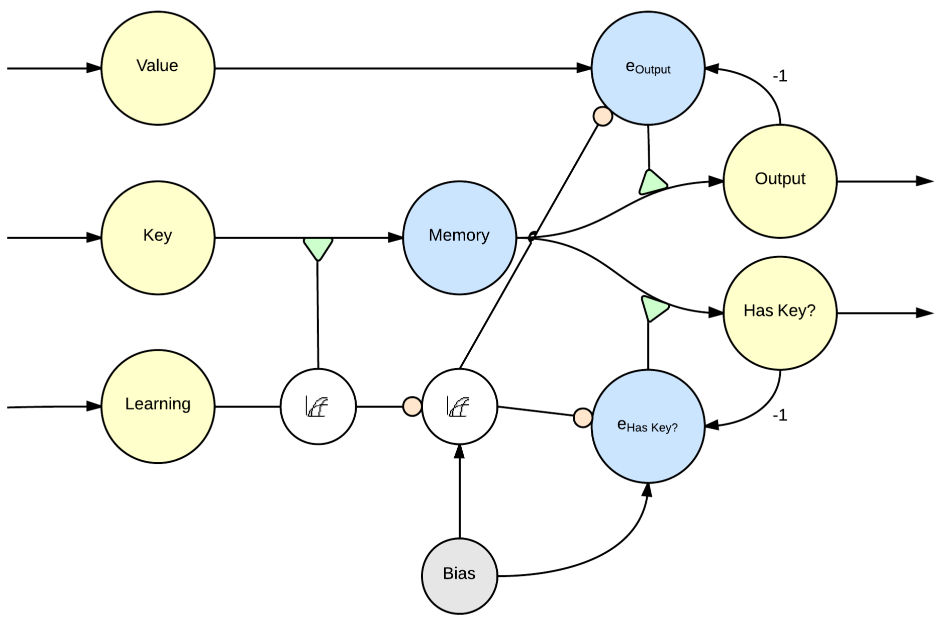

Figure 3: Design of an associative memory in Nengo. Yellow circles are "passthrough nodes", blue circles are NEF ensembles, white circles are scalar $[0, 1]$ ensembles with positive intercepts and preferred direction vectors, orange connections are inhibitory, and green connections are modulatory. The "bias" is a constant scalar input of 1 (see text for details).

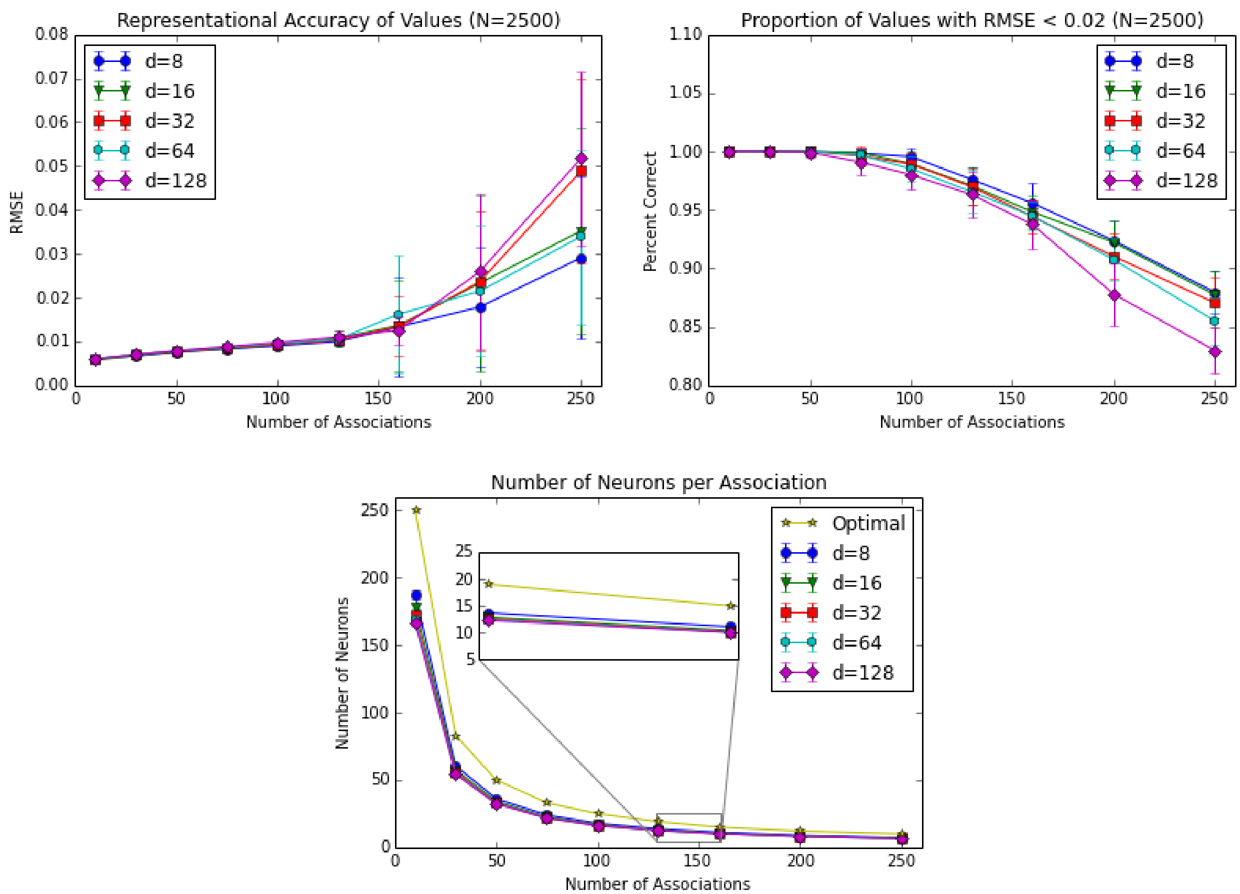

Figure 4: One-dimensional values stored in an associative memory containing 2,500 spiking LIF neurons in its memory ensemble. The vocabulary sizes and dimensions are varied in separate tests. Each test first learns all associations until convergence, and then recalls all values in the same order. The error is measured using a post-synaptic filter of 0.02, from 110ms until 130ms after each key is shown. Every test is repeated 25 times, with the mean and standard deviation reported. (Left) The RMSE over all values. (Right) The percentage of values recalled "correctly" (RMSE < 0.02). (Bottom) The average number of neurons that fire in response to each key (note the standard deviation between trials is close to zero). The theoretical optimal is $2500/n$.

References

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought.* Erlbaum Associates.

Bekolay, T., Kolbeck, C., & Eliasmith, C. (2013). Simultaneous unsupervised and supervised learning of cognitive functions in biologically plausible spiking neural networks. In *Proceedings of the 35th annual conference of the cognitive science society* (pp. 169–174).

Crawford, E., Gingerich, M., & Eliasmith, C. (2013). Biologically plausible, human-scale knowledge representation. In *Proceedings of the 35th annual conference of the cognitive science society.*

Eliasmith, C., & Anderson, C. H. (2003). *Neural engineering: Computation, representation, and dynamics in neurobiological systems.* MIT Press.

Eliasmith, C., Stewart, T. C., Choo, X., Bekolay, T., DeWolf, T., Tang, Y., & Rasmussen, D. (2012). A large-scale model of the functioning brain. *science*, *338*(6111), 1202–1205.

Hassoun, M. H. (1989). Dynamic heteroassociative neural memories. *Neural Networks*, *2*(4), 275–287.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, *79*(8), 2554–2558.

Kohonen, T. (1972, April). Correlation matrix memories. *IEEE Trans. Comput.*, *21*(4), 353–359. doi: 10.1109/TC.1972.5008975

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological cybernetics*, *43*(1), 59–69.

Oja, E. (1989). Neural networks, principal components, and subspaces. *International journal of neural systems*, *1*(01), 61–68.

Pollack, J. B. (1988). Recursive auto-associative memory: devising compositional distributed representations. In *Proceedings of the 10th annual conference of the cognitive science society.*

Stewart, T. C., Tang, Y., & Eliasmith, C. (2011). A biologically realistic cleanup memory: Autoassociation in spiking neurons. *Cognitive Systems Research*, *12*(2), 84–92.

Treves, A., & Rolls, E. T. (1994). Computational analysis of the role of the hippocampus in memory. *Hippocampus*, *4*(3), 374–391.

Willshaw, D. J., Buneman, O. P., & Higgins, L. H. C. (1969). Non-holographic associative memory. *Nature*, *222*, 960–962.

Wu, Y., & Pados, D. A. (2000). A feedforward bidirectional associative memory. *Neural Networks, IEEE Transactions on*, *11*(4), 859–866.