Appendix: Generating new concepts with hybrid neuro-symbolic models

Reuben Feinman (reuben.feinman@nyu.edu)
Center for Neural Science
New York University

Brenden M. Lake (brenden@nyu.edu)
Department of Psychology and Center for Data Science
New York University

A. Model Hyperparameters

Here we review the hyperparameters (HPs) used for each of our models, indicating which HPs were fixed and which were tuned. All neural networks with GMM output layers use 20 mixture components.

Full NS. The Full NS model has 3 submodules: a *location* model, a *stroke* model, and a *termination* model. Each submodule uses a distinct CNN, and each receives an image canvas of size (28,28). The *location* and *termination* models–which return outputs for a single time step–each use a feed-forward CNN architecture inspired by Vinyals et al. (2016). The CNNs consist of a stack of 4 blocks, with each block i including a 3x3 convolution with K_i filters, batch normalization, nonlinear activation f, 2x2 max-pooling, and dropout with rate p. These blocks are followed by a single fully-connected layer with D units, activation f and dropout p. Hyperparameters $\{K_i\}$, f, p and D were selected from tuning. The *stroke* model uses a modified CNN architecture without spatial pooling, designed to convey high-resolution spatial information for visual attention. The CNN returns a feature map of size (64, 14, 14), which is then passed to an LSTM. The LSTM predicts the spline trajectory of the next stroke one offset at a time, attending to different parts of the feature map at each step. The HPs of the CNN were fixed, but the HPs of the LSTM were tuned, including the number of LSTM layers and number of units per layer.

Hierarchical LSTM. The Hierarchical LSTM model has a character-level LSTM backbone and 3 submodules: a *stroke encoder* (BiLSTM), a *location model* (MLP), and a *stroke model* (LSTM). The number of LSTM layers, number of units per layer and dropout rate in the character-level LSTM were selected from tuning, but HPs of all submodules were fixed. The *stroke encoder* is a bidirectional LSTM with a single layer of 256 units. It outputs a fixed-length vector representation of the previous stroke, which is fed to the character LSTM as input. The *location model* is a 2-layer MLP that receives the current hidden state of the character LSTM and outputs a GMM for the next stroke's starting location. The *stroke model* is an LSTM with a single layer of 256 units and outputs a GMM at each time step for the next spline offset.

Baseline LSTM. The Baseline LSTM is a single module. It has L LSTM layers, each with K units and dropout rate p. The values of L, K and p were selected from tuning.

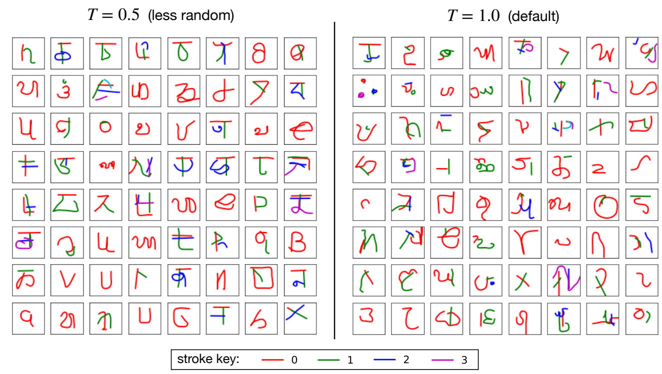

Figure 1: Samples with stroke decomposition. Character samples produced by our Full NS model are shown with stroke decompositions. Samples were produced at two temperature settings (Ha & Eck, 2018, Eq.8), using $T = 1.0$ and $T = 0.5$.

B. Samples with stroke decomposition

In Fig. 1, we show a larger collection of characters from our Full NS model, using color coding to convey the stroke composition of each sample. We produced character samples at two different levels of stochasticity, using a temperature parameter to modify the entropy of the mixture density outputs (Ha & Eck, 2018, Eq.8). Samples are shown for temperature settings $T = 1.0$ and $T = 0.5$.

References

Ha, D., & Eck, D. (2018). A neural representation of sketch drawings. In *ICLR*.

Vinyals, O., Blundell, C., Lillicrap, T., & Wierstra, D. (2016). Matching networks for one shot learning. In *NIPS*.