Assigning Uniqueness to Generative Features for Discrimination

Bonny Banerjee, Juan Gu, Jayanta K. Dutta

Institute for Intelligent Systems, and Dept. of Electrical & Computer Engineering, University of Memphis
206 Engineering Science Bldg., 3815 Central Ave, Memphis, TN 38152, USA
BonnyBanerjee@yahoo.com, jgu@memphis.edu, jkdutta@memphis.edu

The traditional approach to building a brain-inspired neural network based classification system proceeds in two stages. First, the features are learned from sensory data in an unsupervised manner often using a multilayered generative model. The discriminative weights are then learned between the top feature layer and a classification layer in a supervised manner. The classification layer consists of n neurons, each representing a class, which is a highly inefficient representation.

It is more efficient to use a sparse representation whereby a small set of m ($m<<n$) neurons represents each class. Theoretically, n-choose-m classes can then be represented using the n neurons. This is consistent with the hypothesis that the brain uses a sparse representation (Olshausen & Field, 1996). We observe that, while all features learned from the data are necessary for reconstructing different stimuli, only a small subset of features are necessary and sufficient for discriminating a stimulus from those belonging to other classes. That is, if these discriminative or salient features are observed in a stimulus, it can be classified as correctly as when all features are observed. We show how an expected degree of uniqueness or discriminative capability may be assigned to each feature as they are learned in a generative model such that classification may be achieved using a sparse representation. We present a model where the expected uniqueness of a feature is captured by the adaptive threshold of the neuron encoding it. When a stimulus is presented, a neuron is less likely to fire if it encodes a less discriminative feature and vice versa. The firing pattern of the population of neurons encodes the class using a sparse representation.

This model is consistent with recent findings in neuroscience that individual neurons respond to a class of surprise in the stimuli (see for example, Gill et al., 2008; Meyer & Olson, 2011). A surprise is evoked when an unexpected feature occurs in the stimulus. More discriminative features occur less often, are less expected, and therefore evoke greater surprise. Thus, firing pattern of the population of neurons in our model is in response to surprises in the stimuli.

References:

Gill, P., Woolley, S. M. N., Fremouw, T. and Theunissen, F. E. (2008). What's that sound? Auditory area CLM encodes stimulus surprise, not intensity or intensity changes. J. Neurophysiol., 99:2809-2820.

Meyer, T. and Olson, C. R. (2011). Statistical learning of visual transitions in monkey inferotemporal cortex. Proc. Natl. Acad. Sci., 108(48):19401-19406.

Olshausen, B. A. and Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature, 381(6583):607-609.