A Selective Sparse Coding Model with Embedded Attention Mechanism

Qingyong Li, Beijing Jiaotong University, China
Zhiping Shi, Chinese Academy of Sciences, China
Zhongzhi Shi, Chinese Academy of Sciences, China

ABSTRACT

Sparse coding theory demonstrates that the neurons in primary visual cortex form a sparse representation of natural scenes in the viewpoint of statistics, but a typical scene contains many different patterns (corresponding to neurons in cortex) competing for neural representation because of the limited processing capacity of the visual system. We propose an attention-guided sparse coding model (AGSC). This model includes two modules: non-uniform sampling module simulating the process of retina and data-driven attention module based on the response saliency (RS). Our experiment results show that the model notably decreases the number of coefficients that may be activated and retains the main vision information at the same time. It provides a way to improve the coding efficiency for sparse coding model and to achieve good performance in both population sparseness and lifetime sparseness.

Keywords: attention mechanisms; sampling modules; sparse coding theory

INTRODUCTION

Understanding and modeling the functions of the neurons and neural systems are one of the primary goals of cognitive informatics (CI) (Wang, 2002, 2007; Wang & Kinsner, 2006). The computational capabilities and limitations of neurons, and the environment in which the organism lives are two fundamental components driving the evolution and development of such systems. The researchers have broadly investigated them.

The utilization of environmental constraints is most clearly evident in sensory systems, where it has long been assumed that neurons are adapted to the signals to which they are exposed (Simoncelli & Olshausen, 2001). Because not all signals are equally like each other, it is natural to assume that perceptual systems should be able to best process those signals that occur most frequently. Thus, it is the statistical properties of the environment...
that are relevant for sensory process of vision perception (Field, 1987; Simoncelli, 2003).

Efficient coding hypothesis (Barlow, 1961) provides a quantitative relationship between environmental statistics and neural processing. Barlow for the first time hypothesized that the role of early sensory neurons was to remove statistical redundancy in the sensory input. Then, Olshausen and Field (1996) put forward a model, called sparse coding, which made the variables (equivalence of neurons stimulated by the same stimulus in the neurobiology) be activated (i.e., significantly non-zero) only rarely. This model is named SC here. Vinje and Gallant’s (2000) results validated the sparse properties of neural responses under natural stimuli conditions. Afterwards, Bell and Sejnowski (1997) brought forward another sparse coding model based on statistical independence (called SCI) and obtained the same results as Olshausen and Field’s (1996) model. More recent studies can be seen in survey (Simoncelli, 2003).

However, Willmore and Tolhurst (2001) argued that there were two different ways for “sparseness”: population sparseness and lifetime sparseness. Population sparseness describes codes in which few neurons are active at any time and it is utilized in Olshausen and Field’s (1996) sparse coding model; while lifetime sparseness describes codes in which each neuron’s lifetime response distribution has high kurtosis, which is the main contribution in Bell and Sejnowski’s (1997) sparse coding model. In addition, it is proved that lifetime sparseness was uncorrelated with population sparseness. Just as Figure 3a shows the number of variables, which have large values produced by sparse coding model and are possible to be activated, is relatively large compared with the computation capacity of neurons. Though, the kurtosis of every response coefficient is also high. So, how to reduce both population sparseness and lifetime sparseness at the same time to retain the important information as much as possible is a valuable problem in practice.

Vision attention mechanism is an active strategy in information processing procedures of the brain, which has many interesting characteristics such as selectivity and competition. Attention is everywhere in the visual pathway (Britten, 1996). Furthermore, a typical scene within the neuron’s classic receptive field (CRF) contains many different patterns that compete for neural representation because of the limited processing capacity of neurons in the visual system. So, integrating attention mechanism into the sparse coding framework to reduce the population sparseness and to improve the coding efficiency sounds reasonable and essential.

In this article, we extend the sparse coding principle combining the vision attention. We first model the sampling mechanism of the retina by a non-uniform sampling module; then, we implement a bottom-up attention mechanism based on the RS of the sparse coefficient. The diagram is illustrated in Figure 1. This model has two main contributions:

1. modeling the vision attention in the framework of sparse coding and
2. improving the population sparseness of the response coefficient in the same time retaining the most efficient information.

The rest of the article is organized as follows. The second section presents related work. In the third section, a detailed description of the model is given. Experimental results are presented in the fourth section. Conclusions are given the fifth section.

RELATED WORK

In the sparse coding model (Bell & Sejnowski, 1997; Olshausen and Field, 1996), a perceptual system is exposed to a series of small image patches, drawn from one or more large images, just like the CRF of neurons. Imagine that each image patch, represented by the vector \( x \), has been formed by the linear combination of \( N \) basis functions. The basis functions form the columns of a fixed matrix \( A \). The weight of this linear combination is given by a vector \( s \). Each component of this vector has its own associated basis function and represents a response value of a neuron in the vision system. The linear synthesis model is therefore given by:
Related Content

A Robust Facial Feature Tracking Method Based on Optical Flow and Prior Measurement
[www.igi-global.com/article/robust-facial-feature-tracking-method/49693?camid=4v1a](www.igi-global.com/article/robust-facial-feature-tracking-method/49693?camid=4v1a)

Assessing Inference Patterns
(2012). *Relational Thinking Styles and Natural Intelligence: Assessing Inference Patterns for Computational Modeling* (pp. 62-84).
[www.igi-global.com/chapter/assessing-inference-patterns/65042?camid=4v1a](www.igi-global.com/chapter/assessing-inference-patterns/65042?camid=4v1a)

On Concept Algebra: A Denotational Mathematical Structure for Knowledge and Software Modeling
[www.igi-global.com/article/concept-algebra-denotational-mathematical-structure/1558?camid=4v1a](www.igi-global.com/article/concept-algebra-denotational-mathematical-structure/1558?camid=4v1a)
Humans and Machines: Nature of Learning and Learning of Nature
Hélène Hagège, Christopher Dartnell, Éric Martin and Jean Sallantin (2010). Discoveries and Breakthroughs in Cognitive Informatics and Natural Intelligence (pp. 71-92).
www.igi-global.com/chapter/humans-machines-nature-learning-learning/39260?camid=4v1a