Pattern Recognition 60 (2016) 669–680

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Dynamic saliency-driven associative memories based on network potential field

Shuzhi Sam Ge [a,b,*], Mingming Li [a], Tong Heng Lee [a]

[a] Social Robotics Laboratory, Interactive Digital Media Institute and Department of Electrical and Computer Engineering, National University of Singapore, Singapore 119077, Singapore
[b] Center for Robotics, University of Electronic Science and Technology of China, Chengdu 610054, China

ARTICLE INFO

Article history:
Received 26 February 2015
Received in revised form
21 June 2016
Accepted 22 June 2016
Available online 23 June 2016

Keywords:
Associative memory
Recurrent neural network
Visual saliency
Network potential field
Color image retrieval

ABSTRACT

In this paper, we present a framework to construct a general class of recurrent neural networks (RNNs) as associative memories (AMs) for pattern storage and retrieval. Different from the traditional AM models that treat elements of a pattern equally, the proposed framework introduces visual saliency of a target pattern into the AM design process by encoding saliency values into the ellipsoidal basis function (EBF) kernel that calculates the weighted distance between the input and target patterns. Network potential fields (NPFs) are then constructed as the linear combination of EBF/radial basis function (RBF) kernels for auto-associative memories (AAMs) and hetero-associative memories (HAMs), respectively. Sparse and dynamic synapses for both the proposed AAMs and HAMs are determined according to the gradients of the NPFs with high efficiency and without the continuity assumptions of the RNN's activation function, which is usually required by traditional AMs. With the proposed method, the target patterns are assigned as fixed point attractors of the network and the AMs are proven to be able to converge to one of these attractors via Lyapunov analysis. The resulted AAMs and HAMs are demonstrated to have excellent tolerance and robustness to a variety of input noises and corruption via comparative experiments on retrieval and association of 2D color images.

1. Introduction

Associative memories (AMs) are biologically inspired models that are able to correctly store, recall and associate patterns even when the input patterns are noised, corrupted or incomplete [1,2], which have been applied in recognition [3,4] and classification [5,6]. Through an AM, a mapping is established between the input pattern x and the desired output pattern y, which is considered to be an auto-associative memory (AAM) if $x=y$ or a hetero-associative memory (HAM), otherwise.

In 1982, Hopfield showed that a class of RNN (Hopfield network) can be designed to behave as AMs. Generally, a Hopfield network are determined according to the Hebbian rule that encodes the correlation between elements of a memorized pattern into the network's synaptic weights between different neurons. However, the storage capacity of the Hopfield network is known to be very limited [7] and problems of stability exist. To overcome these problems, extensive works have been done by generalizing the original Hopfield network and Hebbian rule [8–10]. Apart from the classical Hopfield network, other types of RNNs were also designed as AMs [11–15]. Most of these works focused on AMs of small binary patterns. Thus, the efficacy and efficiency of these AMs may degrade while handling large and complex patterns such as images, especially when the input patterns are noised and corrupted.

To realize robust images storage and retrieval, AMs for images retrieval has been a hot research topic in recent decades. Decoupled-Voting Hamming Associative Memory Networks was proposed in [16] for binary image. In [17], the design methodology for a class of RNN with linear threshold (LT) neurons was derived based on Hebbian rule and AAM of gray-scale image was implemented. However, perfect recall was not achieved with noised input. Several different RNN models were adopted in [18–20] for storage and retrieval of gray-scale image, exhibiting ability of

Abbreviations: AM, associative memory; EBF, ellipsoidal basis function; NPF, network potential field; AAM, auto-associative memory; HAM, hetero-associative memory; SD-AAM, saliency-driven AAM; SD-AAM-HAM, a HAM that is concatenated after a SD-AAM; SPN, salt and pepper noise; GN, Gaussian noise; SN, speckle noise; NMSE, normalized mean square error; U-AAM, AAM with uniform visual saliency weights; SR, success rate

* Corresponding author at: Social Robotics Laboratory, Interactive Digital Media Institute and Department of Electrical and Computer Engineering, National University of Singapore, Singapore 119077, Singapore.
E-mail address: samge@nus.edu.sg (S.S. Ge).

http://dx.doi.org/10.1016/j.patcog.2016.06.019
0031-3203/

handling random noises in the input pattern. In [21], an RGB color image was decomposed into three separate gray-scale images and stored in three independent AAMs, which were realized using a class of Cohen–Grossberg networks. Instead of using RNN models, dendritic lattice associative memories [22] were developed for gray scale image, which showed a certain degree of noise tolerance to the input patterns. A Class of Sparsely Connected multi-valued morphological associative memories (MAM) was proposed in [23] for auto-association of large color images and recovering part of the noised or incomplete input. In [24], Hetero-associative morphological memories was developed based on four-dimensional storage. Variants of MAMs were proposed in [25,26] by using the **mid** operator, which have dynamic synapses. A Subspace Projection Associative Memories (SPAM) was proposed in [27], which updates the network's synaptic weights iteratively during pattern retrieval process. Quantale-based associative memory (QAM) was introduced in [28] to improve the noise tolerance for color image retrieval on CIELab color space. Besides MAMs, fuzzy associative memories (FAMs) [29,30] were also developed based on fuzzy neural network where input patterns, output patterns, and/or connection weights of FAMs are all fuzzy-valued. The applicability of FAM on image storage and retrieval were demonstrated in [31–33].

Despite that computational models of AMs are motivated by the biological mechanism in human brains, little attention is paid to the fact that human's internal representation of the complex visual is actually very sparse [34]. In other words, although humans are able to perceive the visual world with rich details, only a minute fraction of the perceived scene is stored in the visual memory because its storage capacity is quite limited [35]. However, the sparsity in visual scene representation also enables humans to robustly store and recall visual memory from an input that might be noised, transformed or corrupted. In [8], the role of saliency weights in memory dynamics has been investigated for a class of attractor network, which shows that the saliency weights can affect the landscape of the attractors. However, only simple correlated patterns rather than images are studied and more importantly, the saliency weights for each pattern are assigned without consideration of the pattern's characteristics, making it less suitable to emulate the biological observations discussed above. To address this problem, this paper proposes to utilize techniques of visual saliency computation [36–38] to compute a saliency map of an image pattern and encode them into our memory storage and retrieval process. The more salient a region is, the more important role it plays when recalling the stored pattern. Such saliency-driven AMs are realized on a class of general continuous-time RNN model, which differs from the previous works that require the neuron activation function to be differentiable and Lipchitz continuous as in [18,21] or discontinuous as in [17,13]. Instead, it has no requirement on the continuity of neuron activation function. Specifically, for auto-association, the obtained saliency maps of every patterns to be stored are encoded in the ellipsoidal basis function (EBF) kernels that compute the weighted distance between the input and the stored patterns. For hetero-association, radial basis function (RBF) kernels are used instead. Then, the network potential field (NPF) is constructed as a linear combination of these EBF/RBF kernels. To improve computational efficiency, connections between neurons are restricted to be sparse as only self-feedback of neurons exist. Synaptic weights for these self-connections are determined according to Lyapunov analysis, which guarantees the asymptotic stability of the resulted RNN. Unlike Hopfield network, where the synapses are fully connected and not adjusted after training phase, the synapses computed by NPF are sparse and dynamic, which yield a desirable computational complexity linear to the pattern dimension and are more biologically plausible [39]. To testify the efficacy of the proposed framework, both saliency-driven AAM (SD-AAM) and SD-AAM-HAM (a HAM that is concatenated after a SD-AAM) are implemented to process color image patterns with a variety of noises or corruption. Experiment results demonstrate that the proposed AMs have excellent noise tolerance and are able to rectify corruptions of input patterns. With the above discussions, the major contributions of this paper are highlighted as follows:

- A new design framework of AMs based on a general class of RNNs is proposed by constructing the NPF with the stored patterns that are explicitly assigned as its local minima. Weights of the sparse dynamic synapses are determined with low computational complexity.
- With the proposed framework, visual saliency is introduced to AMs by shaping the EBF kernels during memory storage and retrieval, which can be considered as a biologically motivated automatic feature extraction process.
- The AMs with the designed sparse and dynamic synapses are guaranteed to be asymptotically stable according to Lyapunov analysis and demonstrated to have excellent tolerance of noise and corruption in input patterns.

The rest of the paper is organized as follows. The network model is presented in Section 2. In Section 3, NPFs are constructed for AAM and HAM, respectively. In Section 4, sparse and dynamic synapses are determined according to NPFs and Lyapunov analysis. Section 5 shows experiments and quantitative analysis on retrieval and hetero-association of color images. Further discussion on the proposed framework and some final remarks are given in Section 6.

2. Network model

Consider the following general RNN model

$$\dot{x}_i(t) = f_i(x_i) + \sum_{j=1}^{n} w_{ij} g_j(x_j) + h_i \tag{1}$$

where $x_i \in \mathbb{R}$ is the state variable of the i-th neuron, $f_i(x)$ is a bounded gain function, $g_i(x)$ is an activation function, h_i is a constant value and w_{ij} is the synaptic weight. Models in [17,18,21] can also be described by Eq. (1). The activation functions in [17] are explicitly defined as a discontinuous linear threshold (LT) function $\sigma_i(x) = \max(x, v_{th})$, where v_{th} is a threshold constant. On the other hand, in [18,21], the activation functions are required to be differentiable and Lipshitz continuous. The model presented here is more general in the sense of less restriction on the activation function as g_i is only required to be nonzero, i.e.,

$$g_i(x) \neq 0, \quad \forall\, x \tag{2}$$

To endow the RNN with the abilities of pattern storage and retrieval, the RNN (1) needs to be properly designed. For simplicity, $f_i(x)$, $g_i(x)$ and h_i are considered to be known functions/value. Thus, it is required to appropriately design the synaptic weights w_{ij} such that the given P patterns $\xi^1, \ldots, \xi^P$ can be assigned as the attractors of the resulted network. As an AM, the network should be able to converge to one of these attractors if the initial state (i.e., input to the network) is sufficiently close. For a color image pattern x composed of n pixels, it can be described by information from C individual channels $x_{(1)}, \ldots, x_{(C)} \in \mathbb{R}^n$. Then, they are concatenated as $x = [x_{(1)}^T, \ldots, x_{(C)}^T]^T \in \mathbb{R}^{Cn}$, which is the neural state of the RNN. The resulted network model in Eq. (1) can be rewritten in matrix form as follows:

$$\dot{x}(t) = F(x) + WG(x) + H \tag{3}$$