Fundamental Models of Artificial Intelligence Inspired by Neuroscience and Cognitive Science

Kai Yi
Institute of Artificial Intelligence and Robotics, Xian Jiaotong University, Xian, Shaanxi, P.R.China
National Engineering Laboratory for Visual Information
Processing and Applications, Xi'an Jiaotong University, Xi'an, Shaanxi, P.R.China
yikai2015@stu.xjtu.edu.cn

Abstract

The abstract goes here.

1. Feedforward Neural Networks

Feedforward neural networks, which mean the units in it not form a cycle, are prevailing in plenty of areas. The pioneering model is Multilayer perceptron (MLP) [1], which can be used to solve linear non-separable problems by using a non-linear activation function. Further, the widely used Deep Neural Network (DNN) [2] has achieved human even superhuman performance on several tasks by making the layers deeper and wider accompanied with some strategies (e.g., dropout, stochastic gradient descent). In this section, we will review the basic model of MLP and DNN respectively. Besides, we will discover the neuroscience and cognitive science bases of the models mentioned above.

1.1. Multilayer perceptron

Biological neural network (BNN) consists of a huge amount of neurons. They make up several subsystems with a unique and complex way interacting among. MLP is a typical feedforward artificial neural network inspired by BNN. In this section, we will review the basic model, learning algorithms of MLP.

A typical MLP consists of hierarchical layers similar to the subsystems of human brains. Further, there are three kinds of layers in MLP: input layer, hidden layer(s) and output layer. Each layer is composed of several nodes, which are often called neuron. Compared to BNN, the nodes in every layer is not connected with each other. To monitor the complicated communication mechanism among subsystems, every node in each layer connect to layers in the next layer (i.e., input layer to hidden layer, hidden layer to hidden layer, hidden layer to output layer) utilizes a nonlinear activation function, which can be used to solve several linear non-separable problems such as speech recognition [3] and character recognition [4]. The topology of MLP is the following Figure 1:

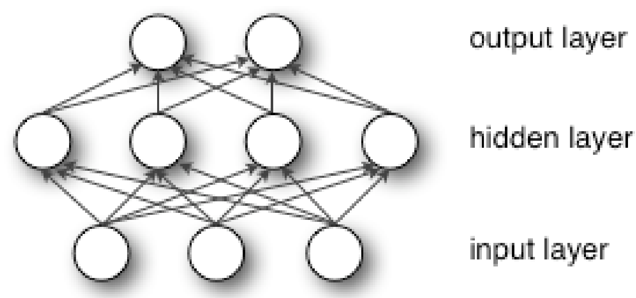

Figure 1. Multilayer perceptron

In the training phase, MLP uses backpropagation, a widely used supervised learning algorithm. MLP just like a high-order non-linear approximate function $f(\cdot)$ that calculate $Y = f(X)$, where X is the vector of input neurons and Y is the output neurons.

Three widely used nonlinear activation function in MLP are sigmoids, adding...

Whether do I need to merge DNN into MLP.

1.2. Deep Neural Networks

Deep neural network is a kind of computational method. It composed of plenty of processing layers that learn and represent the features and the distribution of input data with multiple levels of abstraction (e.g., different depth of feature maps) [5].

Talk about the neuroscience bases of DNN and answer the following question: Whether I can classified DNN to Feebforward Neural Network?

2. Dynamic Neural Network Models

About the definition of Dynamic neural networks.

2.1. Associative Memory Model

Memory is an important function of the biological system, which plays a key role in the process of biological detection and control of the surrounding living environment. Due to the importance of memory in neural system, the study of biological memory can improve the understanding of human brains and promote the design of novel models of artificial intelligence.

Compared to data storage in computers, which is a typical kind of mechanical memory, biological memory has some unique features. They can be briefly classified into four aspects: store and represent information in a distributed way, the relevance between input information and those retrieved, the dynamics of information that is stored or associated, and the correlation with other information processing stage. The second characteristics of biological memory is what we called associative memory [6].

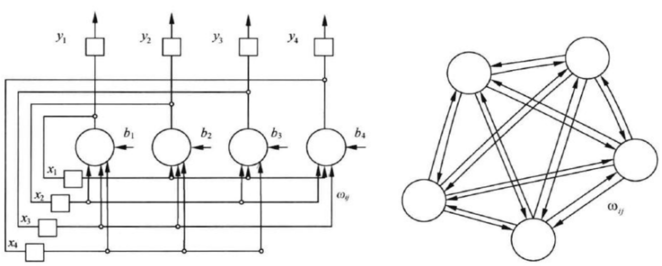

图 9-1 离散 Hopfield 网络　　图 9-2 Hopfield 网络的网状结构

Figure 2. General Structure of Hopfield Neural Network. x_i represents i-th neuron in HNN. · means the multiply of several neurons.

Hopfield Neural Network (HNN) is a popular dynamic neural network. Contrast to Discrete HNNs, which are utilized to solve optimal problems, continuous HNNs (CHNN) are widely used in associative memory [7], [8], the topology of general HNN is shown on Figure 2. The target of HNN is to minimize an energy function (usually minimize to zero), which represents CHNN associates the pattern similar to the input one. The typical definition of energy function of CHNN is the following (1):

$$E = -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}x_i x_j + \sum_{i=1}^{n}\theta_i x_i \tag{1}$$

HNN will meet the convergence when energy function minimize to zero after iterations. It can achieved the global optimal, which is similar to we human brain (local optimal will also occurred, but with small probability).

Currently, the biological-based learning algorithm of HNN is Hebb's learning rule, which assume that "the persistence or repetition of a reverberatory activity tends to induce lasting cellular changes that add to its stability" [9], [10]. Suppose a HNN with N neurons that has memorized M patterns $\xi_\alpha = (\xi_{\alpha 1}, ..., \xi_{\alpha N})^T, \alpha \in [1, M]$, we can get the weights of connections between any two neurons:

$$W = \sum_{\alpha=1}^{M} \xi_a (\xi_a)^T, \;\; w_{ij} = \sum_{\alpha=1}^{M} \xi_{\alpha i}\xi_{\alpha j} \tag{2}$$

where $w_{ij}(i \leq N, j \leq N)$ is the weight between i-th neuron and j-th neuron.

Since the initial neural network energy function is reduced to zero, the maximum memory capacity of the Hopfield network using the Hebb learning rules is smaller.

Since the energy function of the initial neural network is reduced to zero slowly, and the maximum memory capacity of HNN by using the Hebb's rule is very small ($K^{max} \approx 0.14N$). There are a lot of efforts to improve the performance of HNN. Krotov proposed Dense Associative Memory, which can update a novel energy function asynchronously faster than original HNN [8], it also improve the robustness of HNN to some degree [11]. For Hebb's rule has a small memory capacity, pseudo inverse method [12], [13] and perceptron method [14] can improve the capacity the HNN to a large scale. However, the iterations of HNN may come to chaos state, which means it can not associate a fix memory. Although efforts have been made in this field [6], it is still an open question.

2.2. Chaos Neural Networks

A large number of biological experiments have shown that the brain has lots of dynamic behaviors such as bifurcation and chaos [15].

Add: HNN is just one kind of chaos neural network to some degree.

3. Generative Models

Generative model are a branch of fundamental model, which is a statistical process that outputs a synthetic set of data or observations that satisfies a particular distribution. It can produce new data with "imagination". In the field of cognitive science, imagination is very essential for human beings, we can create new art pictures and new concept driven by it.

In this section, we will mainly review two typical generative models: variational autoencoder (VAE) and generate adversarial model (GAN).

3.1. Variational Autoencoder

Review the information decoding and encoding of we human brain.

Currently, information encoding and decoding is of vital importance in building artificial intelligence systems. VAE is a very popular one, which is used to generate new data that similar to input data. It first encodes the input information into a string of hidden vectors through the encoder network, then decodes the implicit vector through the decoder network, and then outputs. Mean square deviation is commonly used as the criterion for evaluating the performance of the encoder. And the weights are trained by the backpropagation algorithm.

Due to the fact that we cannot generate arbitrary data by using traditional autoencoder-decoder network, which means that we can't directly construct hidden vectors that obtained by corresponding input information. VAE is a model that can solve the problem well. The core concept of VAE is to add some restrictions to the coding process, forcing the hidden vectors to obey a normal distribution. What we need to do is giving the network a standard normal distribution of random hidden vectors, so that we can generate the information we want through the decoder, instead of feeding it with raw data and decoding it. The typical structure of VAE is the following:

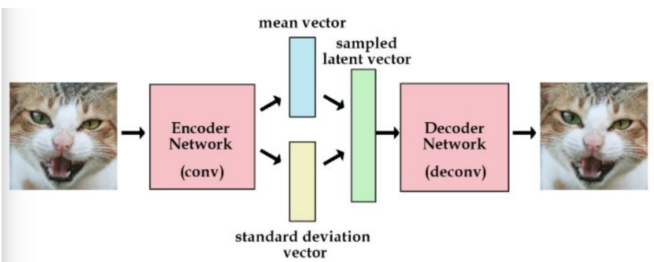

Figure 3. The basic structure of VAE. The input information is encoded through the network to get two vectors, one represents the mean and another represents the standard deviation. Through these two parameters, the hidden vector is synthesized based on Gaussian distribution. Finally, the corresponding information is decoded after the probability distribution is obtained.

The goal of VAE is to get the hidden vector closer to the Gaussian distribution. That is to say that the mean value is 0 and the standard deviation is 1 after the encoding. Generally speaking, Generally speaking, we use KL diverdence to measure the similarity of the two distributions:

$$DKL(P||Q) = \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{q(x)} dx \tag{3}$$

3.2. Generate Adversarial Model

GAN is another popular generative model than is widely used in plenty of fields (e.g., text to image synthesis [16], [17]). It consists of a generative model G that obtain the data distribution, and a discriminative model that evaluate the probability that a sample came from the training data instead of G [18]. The training process of this model is a game indeed, which means G tries its best to maximize the rate of D to make a mistake while D is aimed at discriminate the generated data by D with real training data. The optimal function of D is the following:

$$\max_D E_{\chi\sim P_r}[\log D(\chi)] + E_{\chi\sim P_g}[\log(1 - D(\chi))] \tag{4}$$

For the input x, $D(x) \in [0, 1]$. $D(x)$ is used to evaluate the probability that the given data is real training data. Besides, P_r, P_g represents the distribution of real training data and generated data. The optimal function of G is similar to D. Thus, the optimal function of the model is the following:

$$\min_g \max_D E_{\chi\sim P_r}[\log D(\chi)] + E_{\chi\sim P_g}[\log(1 - D(\chi))] \tag{5}$$

4. Deep Q-Networks

Reinforcement learning is a field that are closer to neuroscience and cognitive science. The learning method is what we called try-and-error, which is popular among the learning procedure of we human[19]. Infants can get feedback and gain knowledge when trying new things, eventually they gain knowledge from this experience.

Talk about the basic structure of reinforcement learning and the basic structure of DQN.

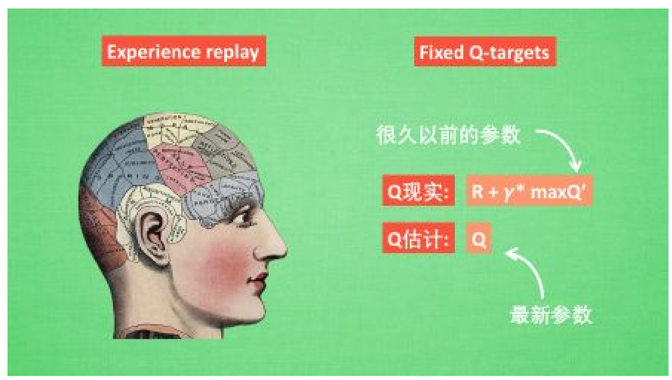

Figure 4. Deep Q-Networks

Briefly speaking

5. Spiking Model

5.1. Spiking Neural Networks

Nowadays, deep neural network achieved great progresses in lots of fields such as image classification [20], [21] and decision-making and control [22], [23]. However, the decoding procedure of DNN is different from biological neural network. The human brain code for information is based on the pulse sequence. When the intensity of the pulse signal reaches a certain peak, that is the action potential, the corresponding neuron produces a certain output. In general, it is a continuous signal instead of a discrete numerical one in a deep neural network.

6. Deep Belief Network

6.1. Restricted Boltzmann Machine

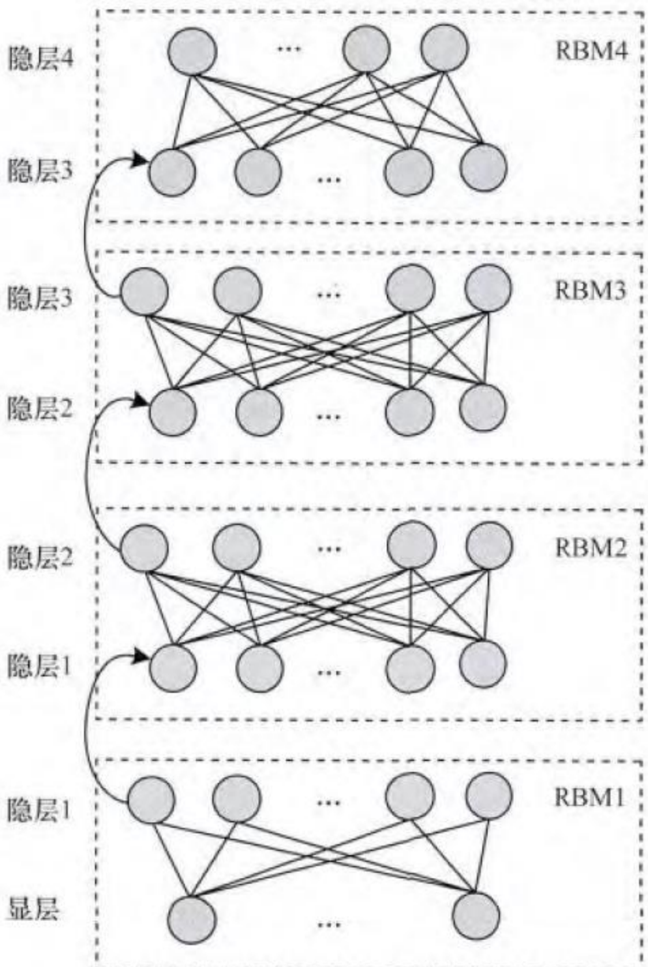

Figure 5. Deep Belief Network

[2][3][4][5] The above figure is the whole training process of DBN. Currently, RBM is widely used in classification [24] / $W = (w_{i,j})$ h_j v_i v_i a_i h_j b_j(v,h)

$$E(v,h) = -\sum_i a_i v_i - \sum_j b_j h_j - \sum_i \sum_j h_j w_{i,j} v_i \quad (6)$$

$$E(v,h) = -a^T v - b^T h - h^T W v \quad (7)$$

References

[1] Frank Rosenblatt. Principles of neurodynamics. perceptrons and the theory of brain mechanisms. Technical report, CORNELL AERONAUTICAL LAB INC BUFFALO NY, 1961.

[2] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436, 2015.

[3] Nelson Morgan and Herve Bourlard. Continuous speech recognition using multilayer perceptrons with hidden markov models. In *Acoustics, Speech, and Signal Processing, 1990. ICASSP-90., 1990 International Conference on*, pages 413–416. IEEE, 1990.

[4] A Khotanzad and JH Lu. Distortion invariant character recognition by a multilayer perceptron and backpropagation learning. In *IEEE International Conference on Neural Networks*, volume 1, pages 625–632, 1988.

[5] Yi Kai, Chen Shitao, Chen Yu, Xia Chao, and Zheng Nanning. Cognition-based deep learning: Progresses and perspectives. In *Internation Conference on Artificial Intelligence Applications and Innovations*, 2018.

[6] Li Yaoyong. Associative memory model: Hopfield neural network and dynamic neural network (chinese). In *PHD thesis of Xi'an Jiaotong University*, 1998.

[7] John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982.

[8] Dmitry Krotov and John J Hopfield. Dense associative memory for pattern recognition. 2016.

[9] Donald Olding Hebb. *The organization of behavior: A neuropsychological theory*. Psychology Press, 2005.

[10] Simon Haykin. Neural networks: a comprehensive foundation, 1999. *Mc Millan, New Jersey*, 2010.

[11] Dmitry Krotov and John J Hopfield. Dense associative memory is robust to adversarial inputs. *arXiv preprint arXiv:1701.00939*, 2017.

[12] Teuvo Kohonen. *Self-organization and associative memory*, volume 8. Springer Science & Business Media, 2012.

[13] Berndt Müller, Joachim Reinhardt, and Michael T Strickland. *Neural networks: an introduction*. Springer Science & Business Media, 2012.

[14] Zong-Ben Xu, Guo-Qing Hu, and Chung-Ping Kwong. Asymmetric hopfield-type networks: theory and applications. *Neural Networks*, 9(3):483–501, 1996.

[15] Kazuyuki Aihara, T Takabe, and M Toyoda. Chaotic neural networks. *Physics letters A*, 144(6-7):333–340, 1990.

[16] Scott Reed, Zeynep Akata, Xinchen Yan, Lajanugen Logeswaran, Bernt Schiele, and Honglak Lee. Generative adversarial text to image synthesis. *arXiv preprint arXiv:1605.05396*, 2016.

[17] Pauline Luc, Camille Couprie, Soumith Chintala, and Jakob Verbeek. Semantic segmentation using adversarial networks. *arXiv preprint arXiv:1611.08408*, 2016.

[18] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[19] Ronald J Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. In *Reinforcement Learning*, pages 5–32. Springer, 1992.

[20] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[21] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[22] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin Riedmiller. Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*, 2013.

[23] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529, 2015.

[24] Hugo Larochelle and Yoshua Bengio. Classification using discriminative restricted boltzmann machines. In *Proceedings of the 25th international conference on Machine learning*, pages 536–543. ACM, 2008.