A Brief Overview of Deep Learning and Memristor

Guojian CHENG¹,a, Xiaoxin AN²,b

¹School of Computer Science, Xi'an Shiyou University Xi'an, Shaanxi, 15029188968, China
²School of Computer Science, Xi'an Shiyou University Xi'an, Shaanxi, 15710497052, China
a,gjcheng@xsyu.edu.cn, b,821663586@qq.com

Abstract. At present, the deep learning has become a main topic in academic and engineering domain. It has been considered to be the general solutions to solve various of classification, clustering, regression and generate tasks in all fields, but it needs higher computing power at the same time. In other words, it needs the new equipment possessing lower power consumption and higher efficiency, which further triggers the development of new hardware and computing technology. A kind of promising solution is an emerging equipment known as memristor, which can be naturally mixed into a feasible computing device to realize neural computing to extend computing power of existing hardware. In this paper, more convenience and practically of using memristor to realize deep neural networks are obtained by summarizing mechanism of deep learning and structure of memristor.

1. INTRODUCTION
In recent years, with the rise of artificial intelligence, the advance of computing resources and the improvement of quantity and quality of available data sets, deep learning have been more and more attention, it has been proved to be an effective tool to solve all kinds of classification, clustering, regression and generated tasks, which are largely attributed to converge to the growth of computing power, the quantities of data and the Internet of things application availability, which requires a higher computing power, low power solution and smaller devices. However, in recent years, the physical limitations of CMOS devices and technology have prompted us to step out of the computing thinking of the CMOS era. A kind of promising solution is an emerging equipment known as memristor[1], which can be naturally blended into a viable computing device to implement neural computing, thus extending the computing power of existing hardware. Therefore, using memristor to implement deep neural network provides more possibilities for the development of deep learning.

2. DEEP LEARNING

2.1. Deep Learning Mechanism
What is deep learning? In a sense, deep learning, or deep neural network, is an implementation of machine learning and a subset of machine learning[2]. It has proved to be an effective tool for solving various categorization, regression, and generation tasks by using tagged or untagged data. In another sense, it can be basically summarized as the following three points: (1) Deep learning is a technology to realize machine learning, and it is an important branch of machine learning, as shown in Figure 1; (2) deep learning originates from the study of artificial neural networks. Its model structure is a neural...
network with multiple hidden layers, as shown in Figure 2; (3) deep learning is to combine low-level features to form more abstract high-level features.

The main advantage of deep learning is that it combines feature extraction and data processing into a single algorithm, thus simplifying human tasks. However, precisely because of this, the process of implementing the algorithm requires very high requirements on computer hardware, which requires the convergence of various emerging devices. Memristor is one of them. At the same time, because of this major advantage, it is believed that in the future, it can be a common solution to various classification, clustering, regression, and generation problems in various fields.

2.2. Deep Neural Network
The simplest form of generalized neural network has the input layer, the hidden layer and the output layer. Deep neural network (DNN) is a neural network with more than one hidden layer, and its corresponding shallow neural network (SNN) has only three layers of feedforward neural network[3]. SNN can be expressed as a nonlinear activation function. By increasing the number of hidden layer neurons and increasing the connections between neurons in the network, DNN represented by highly complex functions is formed. Many studies have shown that DNN is more efficient in calculation than SNN, and the expression capacity of DNN is far greater than SNN. Typical DNNs include Convolutional neural network (CNN), Recurrent neural network (RNN), Long short term memory network (LSTM) and Generative adversarial network (GAN). As shown in Table 1, they all have their own characteristics and different application fields.

| DNNs | Main features | Main applications |
|------|---------------|------------------|
| CNN  | Having convolution layer, pooling layer and full connection layer | Image processing, Object detection |
| RNN  | Combining the feedforward neural network and the feedback neural network | Natural language processing, Stock price forecasting |
| LSTM | Special type of RNN | Natural language processing, Text processing |
| GAN  | Unsupervised learning, including Generative Model and Discriminative Model | Semantic segmentation, Image resolution enhancement |
In the hardware implementation of DNN, the main problem is the scalability of network. When the number of layers is large, the scalability of the network results in a significant increase in power consumption and on-chip area. Therefore, memory devices are considered as a solution to this kind of network. Compared with traditional CMOS or FPGA designs, the implementation of DNN based on memristor ensures the scalability of the design and reduces the area and power consumption on the chip\cite{4}. However, there are specific challenges in memory implementations that require further research and implementation.

3. MEMRISTOR

3.1. Introduction to Memristor
Memristor, known as memory resistor in its full name, was proposed in 1971 by Professor CAI Shaotang at the University of California, Berkeley, as the fourth basic element besides Resistor, Capacitor and Inductor\cite{5}. It represents the relationship between charge and magnetic flux. The main feature of memristor is that they can remember the amount of charge flowing through them, that is, the resistor of the device they are made of depends on the amount of charge flowing through them. Memristor is noted for their nonvolatile ability to convert and store resistor states even at zero power. All the theoretical basis of memristor lies in the dependence between magnetic flux and charge, which is not involved in any other basic element. Figure 3 shows the four basic elements of resistor (R), capacitor (C), inductor (L) and memristor (M) and their relationships.

Memristor is widely used, not only in memory arrays, logic gates, etc, but also in neural networks and complex learning systems\cite{6}. The memristor has small area and low power consumption. Therefore, memristor is a promising solution to the scalability problem of large complex systems such as neural networks. In recent years, more and more attention has been paid to memristor and their applications. In a sense, memristor is a kind of double-ended resistor switch polymoronic memory devices that can be compatible with existing IC technologies. In large-scale simulation, modeling of memristor needs to accurately capture process changes and other non-ideal conditions from actual equipment, so as to ensure the effectiveness of DNN structure design using memristor.

3.2. Memristor Theory
The Memristor is a relatively new device, so the memristor theory is still improving. The following is a brief description of the physical model, mathematical model and characteristics of the memristor.

3.2.1. Memristor Model
As shown in Figure 4, the physical model of memristor is composed of two layers of titanium dioxide film sandwiched between two platinum sheet electrodes. Among them, two kinds of semiconductor
materials, TiO\textsubscript{2} and TiO\textsubscript{2-x}, the former is undoped, while the latter is doped. The specific principle is shown in Figure 5.

As shown in Figure 5, set $D$ as the thickness of titanium dioxide film in the memristor, $w$ as the thickness of doped layer of element, $w$ will change with the change of electric field, when the voltage $U>0$, through the forward charge in the memristor, $w$ will increase. On the contrary, $w$ decreases. In this way, the memristor is similar to a sliding rheostat\cite{7}. The charge flowing through the memristor is equivalent to the slider in the sliding rheostat and can be adjusted freely.

The definition of the mathematical model of memristor is expressed by the following formulas:

According to the circuit theory, the resistance value of the memristor is determined by the amount of charge flowing through it. Therefore, as shown in Formula (1):

$$ R_m(q) = \frac{d \phi(q)}{dq} $$ \hspace{1cm} (1)

Based on the formula (1), the voltage and current relationship of the memristor is shown in Formula (2):

$$ V(t) = R_m \times q(t) \times i(t) $$ \hspace{1cm} (2)

Based on the Formula (2), the resistances of the memristor $R_m$ represent the sum of the resistances of the doped and undoped layers, as shown in Formula (3):

$$ R_m(t) = R_{on} \times x(t) + R_{off} \left(1 - x(t)\right) $$ \hspace{1cm} (3)

Based on the Formula (3), $x(t)$ is shown in Formula (4):

$$ x(t) = \frac{w(t)}{D} \in (0,1) $$ \hspace{1cm} (4)

In the above formulas, $R_{on}$ and $R_{off}$ are the resistances of the memristor when the boundary between the doped layer and the undoped layer moves to $w=D$ (the most right end) and $w=0$ (the most left end), $D$ represents the thickness of the titanium dioxide film in the memristor, and $w$ represents the thickness of the doped layer. In addition, the change rate of $w$ is determined by the charged resistance value and charged ion concentration, as shown in Formula (5):
\[ \frac{dx}{dt} = \frac{\eta \mu R_{on}}{D} \times i(t) \] (5)

\( \mu_v = 10^{-14} \text{m}^2\text{s}^{-1}\text{V}^{-1} \) denotes the average of ion mobility, and \( \eta = \pm 1 \) denotes the polarity of the memristor. It can be seen that in the ideal memristor model, the ion migration is linear, but in reality, the memristor is easily affected by the electric field, and the ion migration is non-linear. To solve this problem, we change the Formula (5) to the following Formula (6):

\[ \frac{dx}{dt} = \frac{\eta \mu R_{on}}{D^2} \times i(t) \times f(x) \] (6)

\( f(x) \) is window function, which is represented by the following Formula (7):

\[ f(x) = 1 - (2x - 1)^{\frac{p}{2}} \] (7)

\( p \) is a positive integer and is the control parameter of the window function. Many studies have shown that the larger \( p \) is, the more ideal the simulated memristor is, and when \( p = 4 \), the memristor is closest to the ideal state.

3.2.2. Characteristics of Memristor
Memristor, as a new kind of hardware, has the irreplaceable advantages of traditional equipment and is also an ideal choice for the next generation of storage devices. Its specific features are as follows:

1. The input and output of memristor are nonlinear, but they are all continuous, so the storage accuracy is theoretically infinite;
2. The memristor, as a basic component, can be directly used in hybrid circuits;
3. The memristor is non-volatile when the charge flows through it, although its internal structure changes, it can be maintained in a new state;
4. The memristor has a small area, low power consumption and is similar to sensor simulation calculation when conducting neural network simulation.

4. DEEP LEARNING AND MEMRISTOR
Over the past few years, rapid advances in electronic, computing and communication technology have led to the increase in data-driven information processing, with neural networks winning because of their simplicity and DNN's effectiveness and accuracy taking the lead. This further triggered the development of emerging hardware and computing technology to achieve larger, more energy-efficient and faster neural networks to solve complex practical problems related to artificial intelligence.

As a new kind of hardware, memristor provides the possibility for the large-scale implementation of neural networks in integrated chips. Compared with traditional CMOS logic, using memristor circuit for neural network simulation has the advantages of small area, low power consumption and similar to sensor simulation calculation. However, on the basis of memristor within DNN implementation, there are some challenges and problems, such as the variability of equipment and the immaturity of technology, the durability of memristor, etc, these will influence the realization of neural network, so the implementation of deep neural network based on the hardware is still a problem to solve, memristor provides the more possibilities for it.

5. CONCLUSIONS
The memristor, as the fourth basic element, is distinguished from the other three basic elements, namely resistor, capacitor and inductor. It is non-volatile and has been widely used in non-volatile storage, big data storage, programmable logic circuit and other fields. In addition to the characteristics of non-volatile, memristor also has the characteristics of low power consumption, small area and high efficiency, in addition, it is also a nonlinear circuit element, precisely because of these characteristics, memristor has been widely used in biomimetic and neural network, nonlinear circuits and other aspects. However, as an important branch of machine learning, deep learning requires high computing power to solve classification, clustering, regression and generation tasks in various fields, which means emerging devices with low power consumption, small area and high efficiency, which is consistent with the
uniqueness of memristor. Many studies have shown that in the traditional neural network circuit design, dozens of transistors, resistors, capacitors and other high-power and large-volume components are needed to simulate a neural unit circuit, which greatly limits the development of neural network. Memristor is a nano-level basic component integrating storage and computing. Its information storage and processing characteristics are very similar to human brain synapses. Therefore, it has become a hot research field to use memristor as the synapses of neurons to construct neural networks. In a word, memristor, as a new memory element, has a wide application prospect and research space. It provides more convenience and practicability for the development of DNN.

REFERENCES
[1] Haowen, T. Research on Design of Memristive Multilayer Neural Network and its Application [D]. University of Electronic Science and Technology of China,2020.
[2] Dianhai, Y. and Tian, W. An Overview of Deep Learning Technology and Platform Development [J]. Artificial Intelligence,2020.
[3] Lili, G. and Baoning, N. Survey of Deep Neural Networks Model Compression [J]. Computer Science and Exploration,2020.
[4] Hui, M., Wei, C. and Jiao, Z., etal. Finite-Time Projective Synchronization of Fractional-Order Memristive Neural Networks with Mixed Time Varying Delays. 2020.
[5] Vishal, G., Giulia, L., Sergio, C. H., etal. Characterisation & modelling of perovskite-based synaptic memristor device[J]. Microelectronics Reliability, 2020, 111.
[6] Yaoyuan, W., Shuang, W., Lei, T., etal. SSM: a high-performance scheme for in situ training of imprecise memristor neural networks[J]. Neurocomputing, 2020, 407.
[7] Bin, W., Fang, F., Wei, W., etal. Modeling of variable resistors based on memristor [J]. Journal of Hefei University of Technology (Natural Science),2020,43(02):188-193.
[8] Jingjing, Y. A Study on Artificial Neuromorphic Devices and Netural Network based on Novel High-Density TiOx Memristors [D]. Anhui University,2020.
[9] Dezhi, G., Yuan, R. and Xianyang J. Design of memristor-COMS hybrid modular inverse circuit [J]. Information technology,2020,44(04):10-16+22.
[10] Xiaoyuan, W., Chenxi, J. and Pengfei, Z. Memristive Digital Logic Circuit Design [J]. Journal of electronics and information technology,2020,42(04):851-861.
[11] Yin, S., Lewis, F., Zhigang, Z., etal. Lagrange Stability and Finite-Time Stabilization of Fuzzy Memristive Neural Networks With Hybrid Time-Varying Delays[J]. IEEE transactions on cybernetics,2020,50(7).