



# **Deep Neural Network and Its Applications**

## **VE370 Project 3**

Bingcheng HU  
516021910219  
December 7, 2018

# Contents

## **Deep Neural Network and Its Applications**

VE370 Project 3

Contents

Abstract

1 Introduction

2 Background

2.1 Biological model and mathematical model of the neuron

2.2 Architectures: Restricted Boltzmann Machine

2.3 Architectures: Deep Belief Network

2.4 Architectures: Deep Convolutional Neural Networks

Convolution and sub-sampling

Pooling

3 Applications

3.1 Speech Recognition

3.2 Computer Vision and Pattern Recognition

3.3 Playing board games

3.4 Predicting game outcomes in Dota 2

4 Topics for future research

4.1 Design a depth model to learn from less training data

4.2 Optimization algorithms for adjusting network parameters

4.3 Unsupervised, semi-supervised and intensive learning methods applied to DNNs in complex systems

5 Discussion

6 Conclusion

References

# Abstract

Artificial intelligence is an area of computer science, it has its own application in almost all fields of science and technology. Since the fast learning algorithm of artificial neural network was proposed in 2006, the deep learning technology has aroused more and more research interest because it overcomes the inherent defect of the traditional algorithm which relies on the artificial design features. Artificial neural network method is also found to be suitable for big data analysis, and has been successfully applied to computer vision, speech recognition, playing board games and predicting the results of Dota2 games. In this paper, we discuss the structures of some widely used artificial neural networks and their practical applications. In this paper, the biological model and mathematical model of neurons are compared, and three kinds of neural network structures are proposed: restricted Boltzmann machine structure, deep belief network structure and deep convolution neural network structure. Finally, a clear reason is given, and the future research topics are listed. **Keywords:** convolution neural network, deep learning, artificial neural network (ANN)

## 1 Introduction

The research of deep-level learning technology has aroused great concern, and a series of exciting research results have been reported in the literature. Since 2009, the ImageNet competition has attracted numerous computer vision research groups from academia and industry. In 2012, a research group led by Hinton won the ImageNet Image Classification Competition [3] by means of deep learning. Hinton's team played for the first time and their performance was 10% better than the second. Both Google and Baidu have updated their image search engines on the basis of the Hinton deep learning architecture and made significant improvements in search accuracy. Baidu also set up (IDL), an institute for deep learning, in 2013 and invited Andrew Ng, an associate professor at Stanford University, to be its chief scientist. In March 2016, Google's deep learning program, known as DeepMind, held a go competition in South Korea between their AI player AlphaGo and one of the world's most powerful players, Lee Se-dol [7].

The purpose of this paper is to review and introduce the deep learning technology and its application in time. It aims to provide readers with the background of different in-depth learning frameworks, as well as the latest developments and achievements in this area.

## 2 Background

### 2.1 Biological model and mathematical model of the neuron

Neural network can point to two kinds, one is biological neural network, the other is artificial neural network. Biological neural network generally refers to the biological brain neurons, cells, contacts and other components of the network, used to generate biological awareness, to help living things to think and act and figure 1 (A) shows its image.

Artificial Neural Network (Artificial Neural Networks,) is also called artificial Neural Network (NNs) or connection Model (Connection Model),. It is an algorithmic mathematical model which imitates the behavior characteristics of animal neural network and processes distributed and parallel information as shown in figure 1 (B). This kind of network depends on the complexity of the system, by adjusting the

connection between a large number of internal nodes, so as to achieve the purpose of processing information.

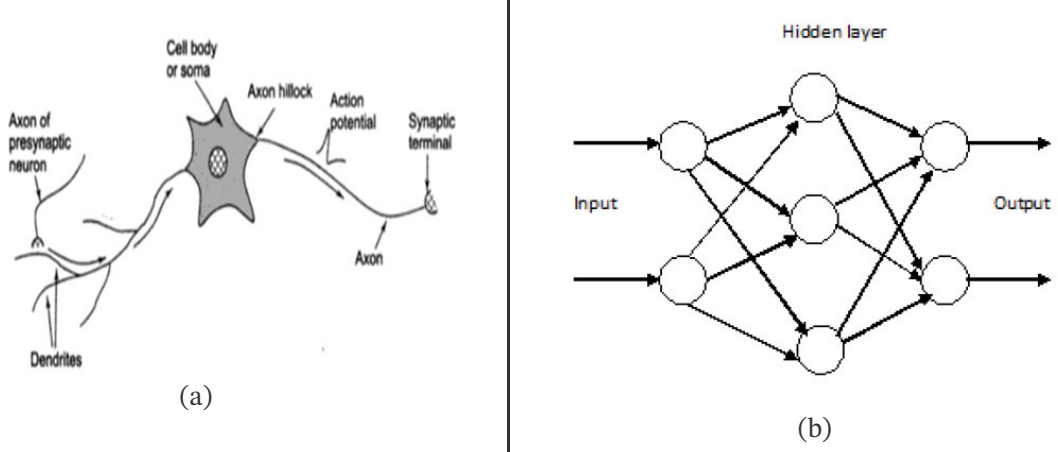


Fig. 1. Biological model of neuron(a) and architecture of Neural Network with hidden layers(b) [5]

## 2.2 Architectures: Restricted Boltzmann Machine

Restricted Boltzmann Machine (RBM) is a kind of generative stochastic neural network, proposed by Hinton and Sejnowski in 1986 [9]. The network consists of some visible elements and some hidden unit, visible variables and hidden variables are binary variables, that is, their states take  $\{0,1\}$ . The entire network is a bipartite graph in which edges exist only between visible units and hidden units, and there is no edge connection between visible units or hidden units.

It should be noted that the training process will be more efficient when using a “gradient based contrast divergence (CD) algorithm”. Hinton develops the CD algorithm for RBM training [9]. Algorithm 1 gives the process of the k-step CD algorithm.

---

**Algorithm 1** k-step contrastive divergence for RBMs

---

**Input:** RBM( $V_1, \dots, V_m, H_1, \dots, H_n$ ), training period  $T$ , learning rate  $\epsilon$

**Output:** The RBM weight matrix  $\omega$ , gradient approximation  $\Delta\omega_{ij}$ ,  $\Delta a_i$  and  $\Delta b_j$  for  $i = 1, \dots, m$ ,  $j = 1, \dots, n$

Initialize  $\omega$  with random values distributed uniformly in  $[0, 1]$ ,  $\Delta\omega_{ij} = \Delta a_i = \Delta b_j = 0$  for  $i = 1, \dots, m$ ,  $j = 1, \dots, n$

For  $\forall t = 1 : T$  do,

    Sample  $v_i^{(t)} \sim P(v_i|h^{(t)})$  when  $i = 1, \dots, m$

    Sample  $h_j^{(t+1)} \sim P(h_j|v^{(t)})$  when  $j = 1, \dots, n$

    For  $i = 1, \dots, m$ ,  $j = 1, \dots, n$  do

$\Delta\omega_{ij} = \Delta\omega_{ij} + \epsilon \times [P(H_j = 1|v^{(0)}) \times v_i^{(0)} - P(H_j = 1|v^{(T)}) \times v_i^{(T)}]$

$\Delta a_i = \Delta a_i + \epsilon \times (v_i^{(0)} - v_i^{(T)})$

$\Delta b_j = \Delta b_j + \epsilon \times [P(H_j = 1|v^{(0)}) - P(H_j = 1|v^{(T)})]$

$\omega = \omega + \epsilon \times \Delta\omega_{ij}$

    End For

End For

---

RBM is an unsupervised learning method. The purpose of unsupervised learning is to fit the training data as much as possible. For a set of input data, it's very difficult to learn if we don't know what distribution it belongs to.

For example, if we know that the data conform to the Gao Si distribution, we can determine the parameters, using the maximum likelihood can easily learn, but if we do not know what distribution, there is no way to start. However, the conclusion of statistical mechanics shows that any probability distribution can be transformed into an energy model. Which makes it possible to fit any data. In Markov random field (MRF), the energy model plays two roles: first, the measure of global solution (objective function); second, the solution with minimum energy is the objective solution. This provides the objective function and solution for the energy model [9].

## 2.3 Architectures: Deep Belief Network

In 2006, Geoffrey Hinton, the father of neural network, gave birth to the artifact, solved the training problem of deep neural network in one stroke, promoted the rapid development of deep learning, and created a new situation of artificial intelligence. In recent years, a lot of intelligent products have emerged in the field of science and technology, which has deeply affected everyone's life. And what is this artifact? That's the Deep Faith Network (Deep Belief Network, or DBN) [14].

Deep belief network (DBN) solves the optimization problem of deep-level neural network by layer-by-layer training, and gives better initial weights to the whole network by layer-by-layer training. As long as the network can be fine-tuned to achieve the optimal solution.

---

**Algorithm 2** Greedy layer-by-layer algorithm for DBN

---

**Input:** Input visible vector  $v_{in}$ , training period  $T$ , learning rate  $\epsilon$ , number of layers  $J$ .

**Output:** Weight matrix  $\omega^i$  of layer  $i, i = 1, 2, \dots, J - 2$ .

Initialize  $\omega$  with random values from 0 to 1,  $h^0 = v_{in}$ ; where  $h^0$  denotes the value of units in the input layer.  $h^i$  represents the units' value of the  $i$ th layer. layer = 1;

for  $\forall t = 1 : T$  do,

    for  $layer < L$ , do gibbs sampling  $h^{layer}$  using  $P(h^{layer} | h^{layer-1})$

    Computing the CD in Algorithm 1,  $\omega(layer)$  is achieved using  $\omega(t + 1) = \omega(t) + \epsilon \times \Delta\omega$

    End for

End for

---

Firstly, the lowest RBM visibility layer is trained, and  $H(0)$  is used as input. Then the values in the visible layer are imported into the hidden layer, and the activation probability  $P(h|\mu)$  of the hidden variables is calculated in the hidden layer. In statistical learning, if we think of the model we need to learn as a high-temperature object, Viewing the process of learning as a process of cooling down to a thermal equilibrium (which in physics usually refers to the stability of temperature in time or space), the energy of the final model will converge to a distribution, fluctuating up and down at the global minimum energy. This process, called simulated annealing, comes from the term "annealing" in metallurgy, in which the material is heated and then cooled by annealing at a certain rate to reduce defects in the lattice. The distribution to which the model energy converges is Boltzmann distribution. As difficult as it sounds, just remember one key point: when energy converges to a minimum, the heat balance stabilizes, that is, when the energy is minimal, the network is the most stable and the network is optimal [3].

## 2.4 Architectures: Deep Convolutional Neural Networks

Convolution neural network is a kind of multi-layer neural network, which is good at dealing with the related machine learning problems of images, especially large images.

Convolution network through a series of methods, successfully reduce the dimension of the image recognition problem with huge amount of data, and finally make it be able to be trained. CNN was first proposed by Yann LeCun and applied to handwritten font recognition (MINST). The network proposed by LeCun is called LeNet, and its network structure is as figure 2.

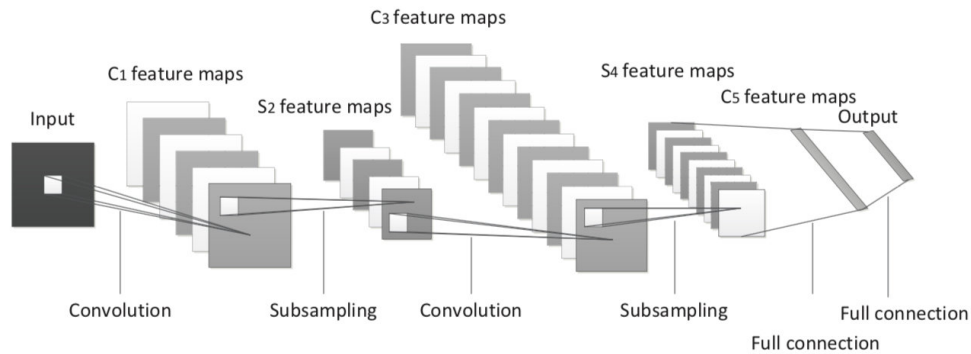


Fig. 2. Schematic structure of CNNs [3]

This is a typical convolution network, which is composed of convolution layer, pooling layer and fully connected layer. Among them, the convolution layer and the pool layer cooperate to form a number of convolution groups, extract the features layer by layer, and finally complete the classification through a number of fully connected layers.

The operation done by the convolution layer can be regarded as inspired by the concept of local receptive field, while the pooling layer is mainly used to reduce the data dimension.

To sum up, CNN simulates feature discrimination through convolution, and reduces the order of magnitude of network parameters through the weight sharing and pooling of convolution. Finally, traditional neural network is used to complete the classification and other tasks.

## Convolution and sub-sampling

Why lower the parameter order of magnitude? It is easy to understand from the following examples.

If we use traditional neural networks to classify an image, then we connect every pixel of the image to the hidden layer node, then for a  $1000 \times 1000$  pixel image, if we have 1m hidden layer units, Then there are altogether  $10^{12}$  parameters, which is obviously unacceptable [7].

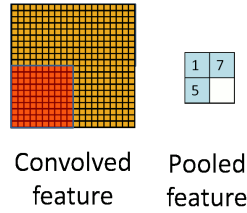
But in CNN, we can greatly reduce the number of parameters, based on the following two assumptions:

1) the lowest-level features are local, that is, we can use filters of the size  $10 \times 10$  to represent low-level features such as edges.

2) the characteristics of different small segments of an image and of small segments of different images are similar, that is to say, we can use the same set of classifiers to describe a variety of different images.

## Pooling

Pooling sounds high and deep, but in simple terms it is downsampling. The process of pooling is shown in the following figure:



The reason for this is that even after convolution, the image is still large (because the convolution kernel is small), so to reduce the dimensions of the data, down-sampling is performed.

This is possible because the statistical properties of the feature can still describe the image even if a lot of data is reduced, and because the data dimensions are reduced, overfitting is effectively avoided [7].

In practical application, pooling can be divided into maximum down sampling (Max-Pooling) and mean down sampling (Mean-Pooling) according to the method of down sampling.

## 3 Applications

With rapid development of computation techniques, the GPU-accelerated computing techniques have been exploited to train CNNs more efficiently. Nowadays, CNNs have already been successfully applied to speech recognition, recommender systems, playing board games and predicting game outcomes in Dota 2.

### 3.1 Speech Recognition

Speech recognition technology refers to the machine automatically converts the content of human voice into text, also known as Automatic Speech Recognition, or ASR technology. Speech recognition is an interdisciplinary and complex subject, which requires knowledge of physiology, acoustics, signal processing, computer science, pattern recognition, linguistics, psychology and other related disciplines [14]. The study of speech recognition is a long and difficult process. Its development can be traced back to the 1950s, when Bell Laboratories first implemented the Audrey English digit recognition system in 1952, which could recognize the pronunciation of a single digit 0 / 9 at that time. And more than 90% accurate to acquaintances. In the same period, MIT and Princeton successively introduced independent word recognition systems for a small number of words. The standard architecture of an ASR system is given in Figure 4.

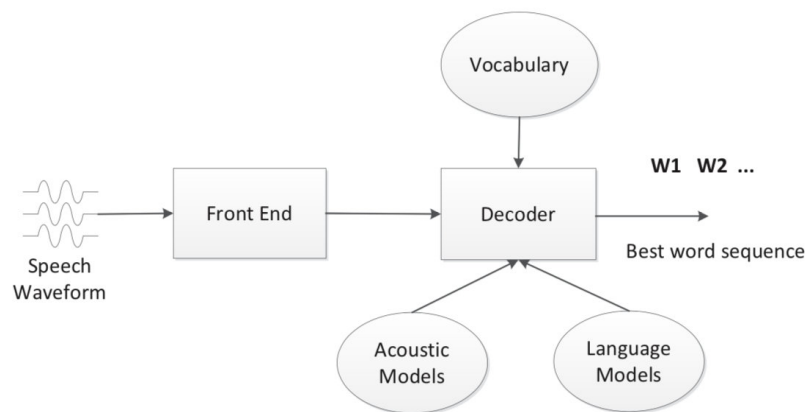


Fig. 4. Speech recognition system architecture [3]

The cumulative probability consists of three parts [11], namely:

Viewing probability: probability per frame and per statu. Transition probability: the probability that each state moves to itself or to the next state. Language probability: probability derived from the laws of language statistics.

Among them, the first two probabilities are obtained from the acoustic model, and the last probability is obtained from the language model. Language models are trained from a large number of texts, which can use the statistical laws of a language itself to help improve the accuracy of recognition. Language model is very important, if the language model is not used, when the state network is large, the result of recognition is basically a mess [11].

## 3.2 Computer Vision and Pattern Recognition

During the past few years, deep learning techniques have achieved tremendous progress in the domains of computer vision and pattern recognition, especially in areas such as object recognition [8]. Deep learning is a hot topic in recent years. Deep learning, inspired by neurology, simulates the process of cognition and expression of human brain, and establishes a logical hierarchy model of implicit relationship in learning data through the function mapping from low-level signals to high-level features. Compared with the ordinary machine learning method of shallow model, the deep learning method has multi-layer structure, and it has better adaptability to big data. As Figure 5 shows, ANN can detect cars, roads, people, ect from the camera [13]. The usage of it is wide.

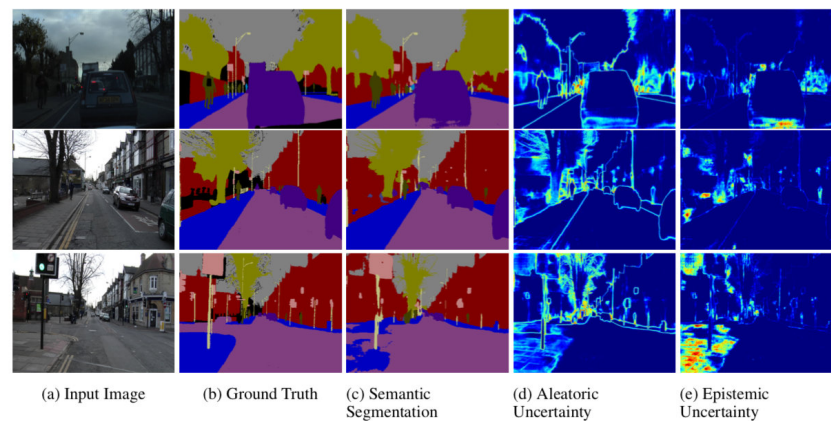


Fig. 5. Computer Vision and Pattern Recognition Usage [13]

DBNs are employed in the computer-aided diagnosis (CAD) systems for early detection of breast cancer [1]. Traditional image processing is only the processing of single or a few digital images, but in today's era of information explosion, the processing of images is more related to the analysis of video (image stream). Each image may have hundreds of thousands of pixels, and if a multi-frame video consists of hundreds or thousands of images, the amount of data is comparable to that of "big data" in other industries. Traditional image object classification and detection algorithms and strategies are difficult to meet the requirements of processing efficiency, performance and intelligence proposed by big data in image and video. Deep learning constructs the mapping from low-level signals to high-level semantics by simulating the hierarchical structure similar to human brain, in order to realize the hierarchical feature expression of data and has the powerful ability of visual information processing. Therefore, in the field of machine vision, Convolution neural network (CNN), the representative of deep learning, has been widely used.



### 3.3 Playing board games

Tic-tac-toe (also known as noughts and crosses or Xs and Os) is a pen and paper game for two players  $\times$  and  $\circ$ , which alternately mark space in a  $3 \times 3$  grid. A player who successfully placed three markers on a horizontal, vertical or diagonal line wins the game [12].

In this experiment, each square in the large well was divided into 9 small squares and one small well. Initially, the first player selects 1 large square and places  $\times$  or  $\circ$  in one of the 9 small squares of the large square. Then the next player should choose a large square whose relative position is the same as the relative positive value of the small square of the well. The game is like this rule [6].

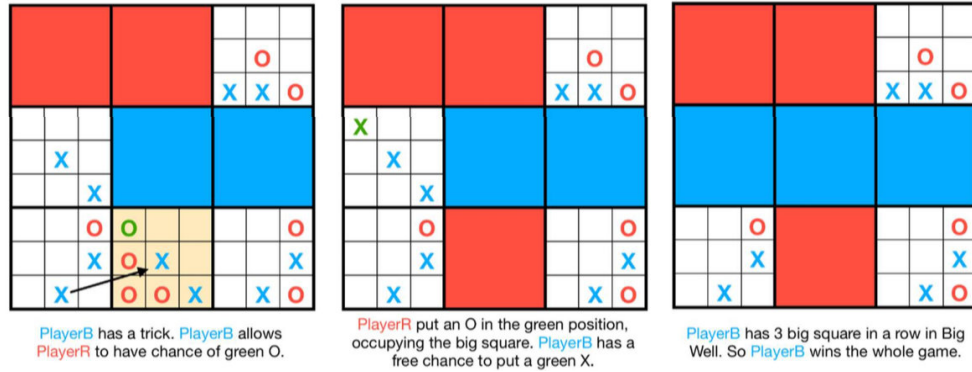


Fig. 6. Blue player wins a Tic-Tac-Toe in Big Well, and the whole game [6]

As shown in Figure 6, by entering the data set generated by the Monte Carlo tree search, the algorithm can automatically improve its board game skills. Because the architecture of the neural network is simple, the training efficiency is not very good. If more versions are developed, a better architecture can lead to better performance [6].

### 3.4 Predicting game outcomes in Dota 2

Dota 2 is “an online strategy game, played in a five versus five format. Its multitude of selectable characters, each with a unique set of abilities and spells, causes every new match to be different from the last and picking the right characters can ultimately decide whether a team wins or loses a game [4].”

In order for an artificial neural network to be useful, it needs to be able to learn. It does this using a technique called back-propagation, which allows it to quickly update weights between neurons in the ANN. [8] the back-propagation algorithm uses the gradient partial E of the loss function E in partial w relative to any weight w. [4] after each epoch, the weight is modified to minimize the mean square error between the prediction result of neural network and the actual target value.

Every ANN was trained for 5000 epochs using the training data. After each step of 1000 epochs we noted the current Mean Squared Error for the training set and tested the ANN. And the testing result is shown as Figure 7.

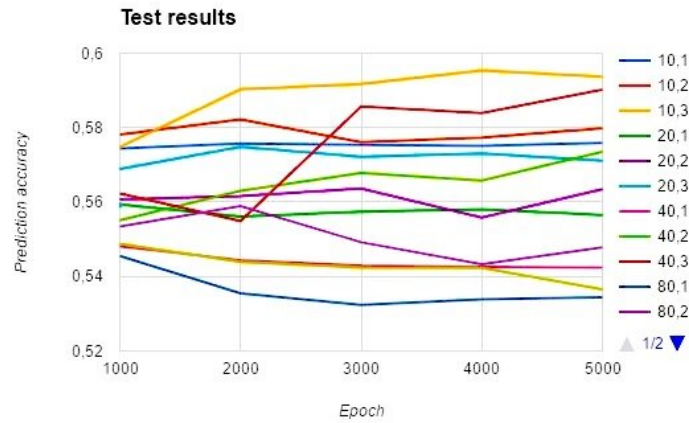


Fig. 7. Testing results after training for different models.[4]

The results show that all models have a prediction rate above 50%. The average accuracy for the models ranges between 53.44% and 59.54%.

## 4 Topics for future research

### 4.1 Design a depth model to learn from less training data

When there is only limited training data, a more powerful model is needed to enhance learning. Therefore, how to learn from less training data, especially in speech and visual recognition systems, how to design a deep-level model is particularly important.

### 4.2 Optimization algorithms for adjusting network parameters

How to adjust the parameters of machine learning algorithms is a new topic in computer science. In DNN, a large number of parameters need to be adjusted. In addition, with the increase of the number of hidden nodes, the algorithm is more likely to fall into local optimum. [8]

### 4.3 Unsupervised, semi-supervised and intensive learning methods applied to DNNs in complex systems

Deep-level learning techniques have not achieved satisfactory results in natural language processing (NLP). With the development of deep unsupervised learning and deep reinforcement learning, there are more options for training DNNs in complex systems.

## 5 Discussion

Deep Neural Network often outperforms humans in a number of platforms and solutions. In the concept of a wide range of areas, machine learning is to enable intelligent devices to simulate human mechanical movement, reasoning and problem-solving methods to achieve the goal of the task and because of the excellent performance of machine learning, many aspects have begun to replace human performance. For example, AlphaGo, launched by Google's deep learning technology division, has

successfully defeated China's go player, Ke Jie, who is ranked No. 1 in the world, while in other programs, such as self-driving and image recognition platforms, The reliability and accuracy of its vision of the environment has surpassed human performance in this field.

There are many things that our current Deep Neural Network can do, such as when listening to music or watching videos, they are recommended according to their preferences, and by tracking past shopping and browsing habits. Shopping websites recommend products or services. The application in these platforms is merely a standardized push, and its intelligence is merely a programmed process formed after the education of the human being.

And Deep Neural Network devices are not necessarily the same as in the movies, so that everyone has a dedicated robot service. In fact, we do not need a machine, in the service has now been Siri, Alexa, Cortana and other products are considered very smart, but it is still not really intelligent.

Deep Neural Network is an application of artificial intelligence, which creates artificial intelligence in a statistical and data-driven way to help computer programs improve performance and accomplish learning tasks. Machine learning depends heavily on data, and the quality of the data or the process of creating it is critical to the success or failure of Deep Neural Network.

## **6 Conclusion**

In 2016, Google artificial Intelligence "AlphaGo" and South Korea's professional go player Lee se-dol launched a "man-machine war", Lee se-dol's defeat made people once again amazed at the power of artificial intelligence. With the maturity of technology, machine learning is widely used in education, medical treatment, image recognition, speech recognition and other fields. In the future, machine learning may touch a variety of industries and change people's daily lives, which also raises widespread concerns about the security of artificial intelligence. Will artificial intelligence be the last invention of mankind? As machines become more human-like and capable of accomplishing human tasks, does this mean that they will take away human jobs? Perhaps for a machine, deep neural network is a more appropriate approach to explore itself.

# References

- [1] Kubat, M. (2015). Artificial neural networks. In *An Introduction to Machine Learning* (pp. 91-111). Springer, Cham.
- [2] Klerfors, D., & Huston, T. L. (1998). Artificial neural networks. *St. Louis University, St. Louis, Mo.*
- [3] Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., & Alsaadi, F. E. (2017). A survey of deep neural network architectures and their applications. *Neurocomputing*, 234, 11-26.
- [4] Widin, V., & Adler, J. (2017). On using Artificial Neural Network models to predict game outcomes in Dota 2.
- [5] Yadav, A., & Sahu, K. (2017). WIND FORECASTING USING ARTIFICIAL NEURAL NETWORKS: A SURVEY AND TAXONOMY. *International Journal of Research In Science & Engineering*, 3.
- [6] Chen, W. (2017). Using Neural Network and Monte-Carlo Tree Search to Play the Game TEN.
- [7] Deng, L. (2012). Three classes of deep learning architectures and their applications: a tutorial survey. *APSIPA transactions on signal and information processing*.
- [8] Zeng, N., Wang, Z., Zhang, H., & Alsaadi, F. E. (2016). A novel switching delayed PSO algorithm for estimating unknown parameters of lateral flow immunoassay. *Cognitive Computation*, 8(2), 143-152.
- [H. Chen] Chen, H., Liang, J., & Wang, Z. (2016). Pinning controllability of autonomous Boolean control networks. *Science China Information Sciences*, 59(7), 070107.
- [9] Hinton, G. E., Osindero, S., & Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. *Neural computation*, 18(7), 1527-1554.
- [10] Hu, J., Chen, D., & Du, J. (2014). State estimation for a class of discrete nonlinear systems with randomly occurring uncertainties and distributed sensor delays. *International Journal of General Systems*, 43(3-4), 387-401.
- [11] Jaitly, N., & Hinton, G. (2011, May). Learning a better representation of speech soundwaves using restricted boltzmann machines. In *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on* (pp. 5884-5887). IEEE.
- [12] Kaplan, E. (1999). U.S. Patent No. 5,927,714. Washington, DC: U.S. Patent and Trademark Office.
- [13] Kendall, A., & Gal, Y. (2017). What uncertainties do we need in bayesian deep learning for computer vision?. In *Advances in neural information processing systems* (pp. 5574-5584).
- [14] Yu, D., & Deng, L. (2011). Deep learning and its applications to signal and information processing [exploratory dsp]. *IEEE Signal Processing Magazine*, 28(1), 145-154.