

Analysis of the status quo and development trend of artificial intelligence AI industry

Hao Liu ^a, Qian Zhang ^b

College of Mechanical and Electronic Engineering, Shandong University of Science and
Technology, Taian, Shandong, China

^abrookfrank@163.com, ^b812193040@qq.com

Abstract: 2016 is the first year of artificial intelligence, and the victory of Alpha Go has ignited the enthusiasm of the whole people. Nowadays, artificial intelligence is already ubiquitous, open your mobile phone, and each app is artificial intelligence: Baidu's search bar will always jump out of the terms you want; Taobao's home page recommendation is always the baby you want; The recommended store is always where you want to go. Depth today Learning to enter a blowout era, it is a smart brain, will humans be replaced in the future? In the current flood of new ideas, it is said that 70% of manufacturing workers will be laid off, "mechanical manual labor" will be replaced; 90% of doctors, programmers and editors will be unemployed, "repeatable experience judgment" Being replaced; college students' English 4th and 6th grades are not required to be tested. It is not necessary to learn foreign languages. Mobile translation software will do better than you. Human beings will be AI Replace it? This article will give you the answer.

Keywords: artificial intelligence, deep learning, development trend, analysis.

1. INTRODUCTION

There are hundreds to hundreds of billions of neurons in the human brain, and each neuron is connected to other neurons through thousands of "synapses", forming a super-large and complex network of neurons for distribution and concurrency. The way to conduct signals is equivalent to ultra-large-scale parallel computing (Parallel Computing). So although a single neuron transmits signals at a very slow rate (a level of 100 meters per second, much lower than the computer's CPU), this ultra-large-scale parallel computing structure still makes the human brain far beyond the computer, becoming the world so far. The most powerful information processing system.

2. THE EMERGENCE OF NEURAL NETWORKS TO POWER THE EMERGENCE OF ARTIFICIAL INTELLIGENCE

Artificial neural network algorithm simulates biological neural network and is a kind of pattern matching algorithm. Usually used to solve classification and regression problems. Important artificial neural network algorithms include: Perceptron Neural Network, Back Propagation, Hopfield Network, and Self-Organizing Map (SOM).

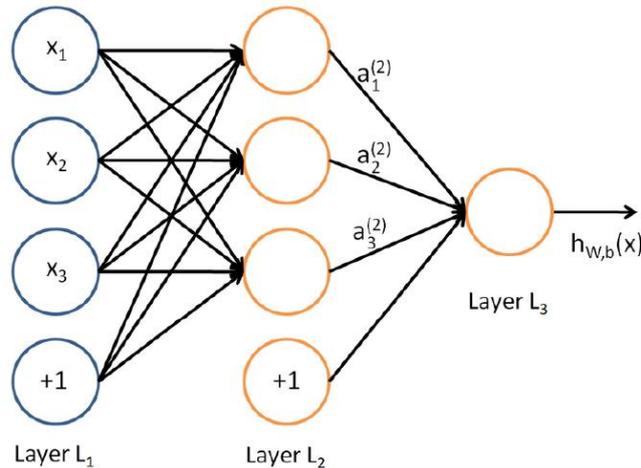


Figure 1: Schematic diagram of an artificial neural network

3. FURTHER DEVELOPMENT OF ARTIFICIAL NEURAL NETWORK BY DEEP LEARNING ALGORITHM

Baidu began to develop deep learning, which has aroused widespread concern in China. Today, as computing power becomes more and more developed, deep learning attempts to build larger and more complex neural networks. Many deep learning algorithms are semi-supervised learning algorithms that deal with large data sets with a small amount of unidentified data. With the support of mathematical models, deep learning has made great progress in image, sound and semantic recognition, especially in the field of image and sound, which greatly improves the recognition rate compared with traditional algorithms. At present, in less than 10 years, deep learning has brought about revolutionary progress in the fields of vision and voice, and detonated the new wave of artificial intelligence.

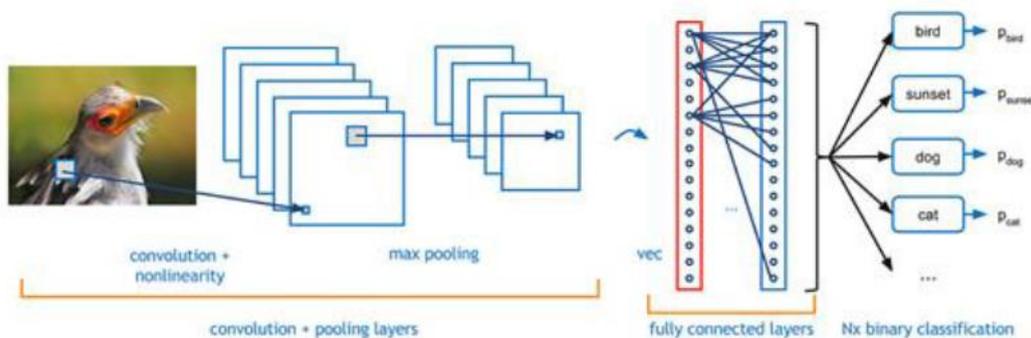


Figure 2: Basically complete deep learning network

4. INCREASED COMPUTING POWER AND STORAGE CAPACITY, PROVIDING THE FOUNDATION FOR THE ARRIVAL OF THE THIRD WAVE OF ARTIFICIAL INTELLIGENCE

In 1946, the computer officially appeared in people's field of vision. It has a history of 70 years, from 5,000 times/second of addition ability to the current peak calculation speed of 559 million times/second; from simple scientific calculation to various fields. The application of data processing; from single-machine processing to global network interconnection and interoperability; from manual connection-driven computing to the birth of the current intelligent brain; the continuous improvement of computer computing capabilities, providing a physical hardware foundation for the arrival of the era of artificial intelligence .

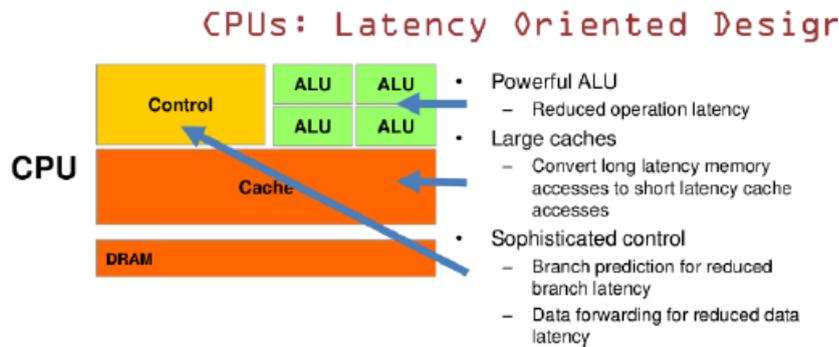


Figure 3: CPU Architecture

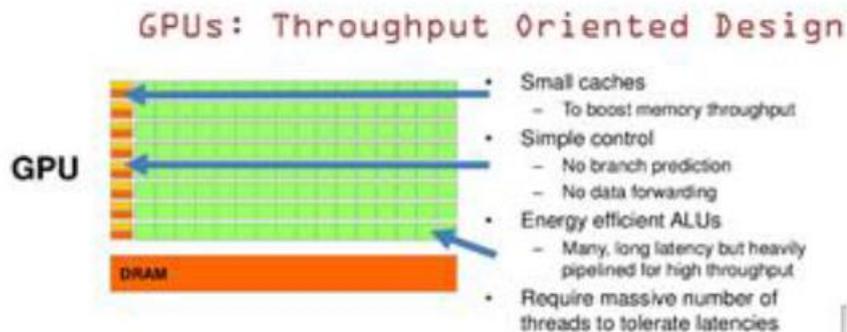


Figure 4: GPU Architecture

The CPU and GPU architectures vary widely, with green being the computing unit, orange being the storage unit, and yellow being the control unit. There are many CPU function modules, which can adapt to complex computing environments; GPU composition is relatively simple, and current stream processors and memory controllers occupy most of the transistors. Compared to the CPU, the GPU's computing speed has a qualitative leap, and it has amazing ability to handle floating-point operations. The emergence of GPUs in 2005 greatly improved the efficiency of computing and led to the success of unsupervised learning technology (one of the techniques involved in deep learning). Moore's Law is not a rule to predict the improvement of CPU performance, but to predict the improvement of semiconductor technology. The law is mainly the integration of transistors. Now CPU technology is slower than Moore's Law, and GPUs (graphics processors on video cards) are running faster than Moore's Law, and their

performance doubles every six months. In 2006, the GPU's computing power was much higher than the CPU's computing power when the GPU and CPU prices were comparable. It can be seen that GPU parallel computing has become a future trend.

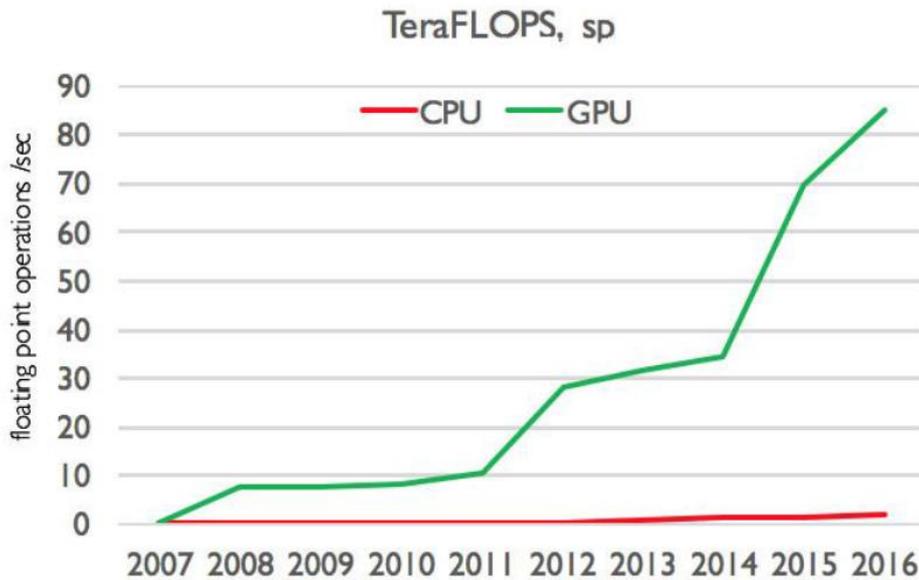


Figure 5: Comparison of CPU and GPU computing power

Parallel computing refers to the simultaneous processing of multiple tasks or multiple instructions, or multiple data items. The computer system that completes this processing is called a parallel computer system, which organizes multiple processors (may be several, dozens, thousands, tens of thousands, etc.) in a certain way through a network connection. . Applying Moore's Law to the growth of chip integration, we can conclude that today means that approximately 50% of components can be placed in a single chip per year. Through the exponential growth of integration, the computational cost is continuously compressed.

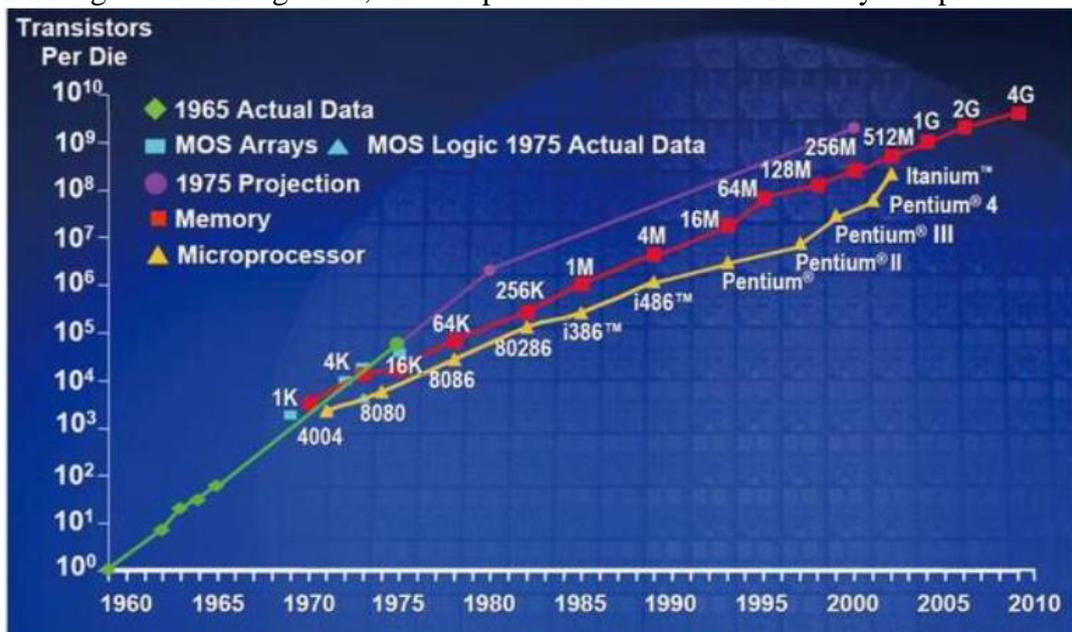


Figure 6: Moore's Law-driven industry

In terms of the storage industry, we can see the evolution of the storage industry, from the initial SAN and NAS storage to today's cloud storage.

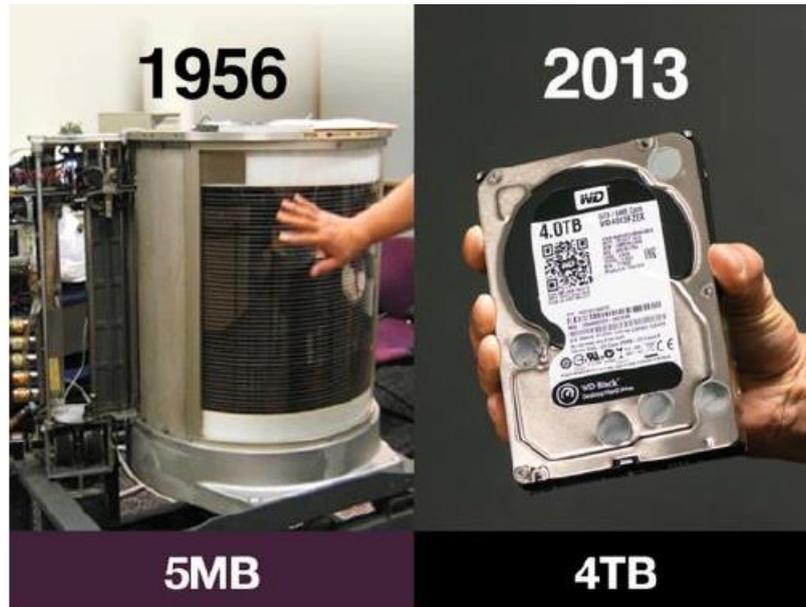


Figure 7: Mechanical Hard Drive: Capacity increased by 1 million times



Figure 8: Server disk storage: 360,000 times improvement

In 1956, the world's first hard disk drive RAMAC 350 appeared, it looks like an air purifier, weighing one ton, storage capacity is only 5MB, then the price is more than 5,000 US dollars (times 13,000 yuan), It can be described as sky-high price. Today, desktop hard drives are typically 3.5 inches in size and can be up to 4TB in a single block. At the same time, hard disk technology is still evolving. In 2020, hard disk using magnetic recording technology (HAMR) can achieve 60TB capacity. The first generation of servers was only used for 256MB of RAM and 2GB of hard disk space, and 22 years later, it evolved to 128GB of RAM and 720TB of hard disk space, achieving a 360,000-fold increase. Of course, the evolution of storage mechanisms and speed cannot be ignored.

The cloud storage system developed based on cloud computing not only enables people to link to the cloud for convenient storage at any time and place, but also through a networkable device. In addition to bringing great convenience to people's lives, cloud storage also greatly reduces the use of mobile storage devices and reduces the cost of enterprises; it can continue to expand in a timely manner based on the number of users and space, and will not affect front-end customers; Real-time synchronization effectively avoids the problem of loss of damage caused by media storage data and improves security.

With the constant confirmation of Moore's Law, the computer computing performance has been greatly improved, coupled with the ever-expanding storage space and the ever-decreasing storage cost, laying the hardware foundation for the rapid development of artificial intelligence.

5. THE DEMAND FROM ALL WALKS OF LIFE CONTINUES TO RISE, ESCORTING THE PERFECTION OF ARTIFICIAL INTELLIGENCE

Nowadays, human demand for artificial intelligence is increasing: in industrial manufacturing, a large number of robots can improve manufacturing efficiency, reduce the rate of product defects, and more importantly, save labor costs; in the field of security, through Video surveillance, face recognition, crowd monitoring and other technologies provide security for the public; in terms of medical health, through the massive data comparison to assist doctors for diagnosis, automatic reading, etc.; in intelligent driving, the accuracy of road sign recognition continues to increase, The combination of image and motion sensors with global positioning systems greatly reduces costs and increases overall safety.

In the short-term, artificial intelligence will achieve greater development in the financial sector; in the medium term, with the continuous improvement of big data technology, massive data accumulation will promote artificial intelligence to achieve new breakthroughs in medical treatment; in the long run, artificial intelligence The ultimate point is that unmanned, computing power, massive data, algorithms and decision-making, and sensor data collection are complete, humans can gradually achieve all-weather, fully automated driverless.

5.1 Universal object recognition is a symbol of the era of true intelligence

In 2012, Google X Labs developed a neural network system with autonomous learning capabilities. It can find pictures of cat faces from pictures without external prompts. At first glance, this is very similar to face recognition in specific object recognition, but in fact, traditional face recognition is to tell the computer what the face should look like when learning the computer, or it can be understood as when training a computer, the data given is pre-labeled. In cat face recognition, the researcher did not tell the computer what the cat face should look like before, but let the computer learn the way through the neural network to mark the image of the feature when dealing with countless pictures. With this technology, different types of objects can be identified in one photo.

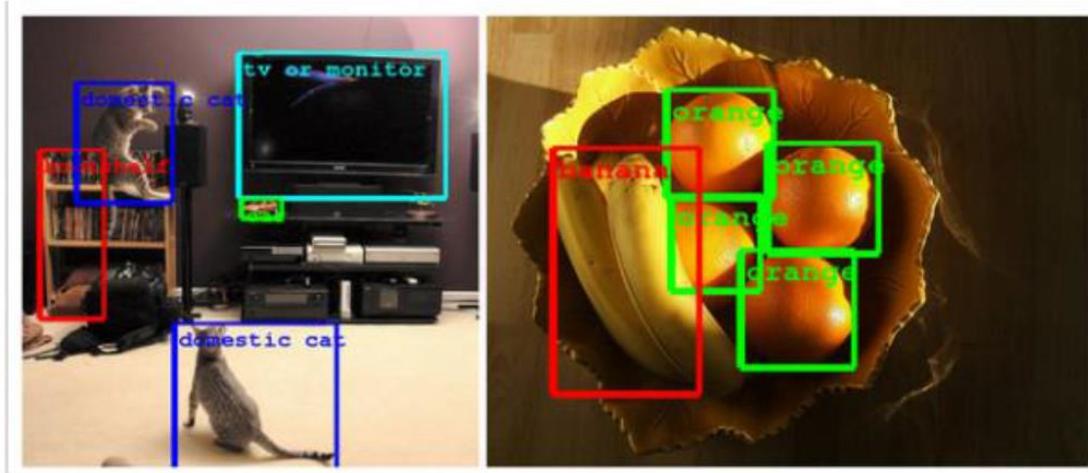


Figure 9: Google can identify various objects on a photo

In order to allow computers to recognize various objects in nature, Google has developed a basic framework called Dist Belief. Using it, Google can call a huge cluster of computers, use billions of parameters for deep learning, and tag different similar objects. Once the computer is again exposed to a similar object, it can recognize that the object is similar to the previous type of tagged object.

Cat face recognition is just a simple demonstration of Google's in-depth learning application. Its core is to try to make the computer recognize the unmarked data on its own. After all, most of the data on the network is unlabeled. What Google is more interested in is the application of this technology in other fields, such as speech recognition, natural language modeling and other fields.

5.2 AI application in the field of cloud computing

Google, which is at the forefront of artificial intelligence, has been working on a thinner, greener cloud computing data platform. Back in 2014, Google minimized energy losses by installing intelligent temperature and lighting controls and using advanced cooling technology instead of mechanical coolers, making its data center power consumption 50% lower than the global data center average. And compared to itself, Google's current data processing performance is 3.5 times five years ago, and energy consumption remains at the original level.

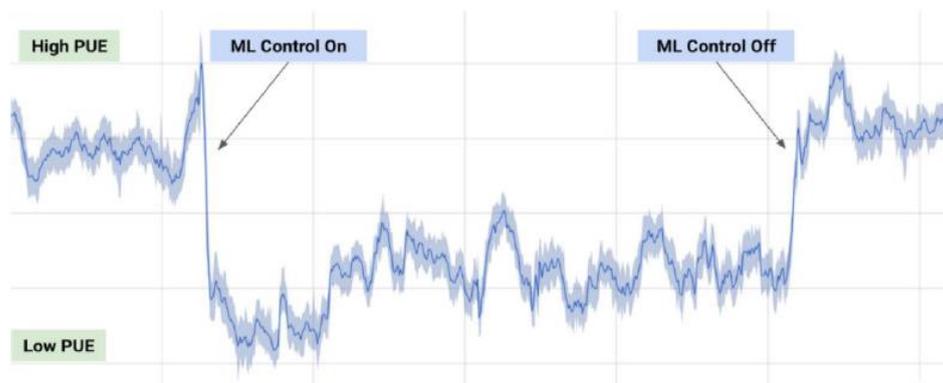


Figure 10: Machine Learning Test Results for Data Center PUE

Today, Google with DeepMind is largely at the forefront of the world. DeepMind will apply the enhanced learning neural network technology to the energy control of cloud computing data centers. By acquiring a large amount of historical data collected by sensors in the data center (such as temperature, power, pump speed, set point, etc.), the first average UE in the future. The (Power Usage Effectiveness) value trains the neural network system. PUE is the ratio of total building energy use to IT energy use and is a standard measure of data center energy efficiency, and each instrument can be affected by dozens of variables. Adjust the model and parameters through continuous simulation to bring it closer to the most accurate predicted configuration and improve the actual performance of the facility. The team trained two additional deep neural network sets to predict the temperature and pressure of the data center in the next hour, simulating the recommended behavior from the PUE model.

Through 18 months of model development and testing, DeepMind and Google Cloud's R&D team successfully saved 40% of the cooling energy and 15% of the total energy consumption in the data center. One of the pilots has reached the lowest point of PUE. Technology may be used to increase power conversion efficiency, reduce energy and water consumption in semiconductor production, or help increase production at production facilities. Machine learning saves energy and reduces carbon emissions in the data center.

DeepMind and the Google Cloud Computing team plan to open up the results to benefit the world's data centers, factories and large buildings to create a greener world.

6. MIT TECH REVIEW 2018 ARTIFICIAL INTELLIGENCE 5 MAJOR TREND FORECAST

Trend 1: Positive reinforcement learning (Positive reinforcement)

The century war between AlphaGo and Li Shishi is a milestone in the field of artificial intelligence, especially deep intensive learning technology. The ultimate goal of artificial intelligence is to imitate the thinking operations of the human brain, and the inspiration for intensive learning comes from the way animals learn. Animals can learn positive or negative outcomes caused by certain behaviors (a positive or negative outcome). In this way, the computer can solve the maze problem by trial and error, and associate the positive result, "going out of the maze," with the behavior that led to the result. This allows the computer to learn without specific instructions or explicit examples.

In fact, the theory of reinforcement learning has existed for decades, but through the combination with large-scale deep neural networks, we have truly acquired the capabilities needed to solve complex problems (such as Go). Through unremitting training and testing, as well as analysis of previous games, AlphaGo is able to find out how to play chess with professional players.

Trend 2: Dueling neural networks

The authoritative artificial intelligence conference held in Barcelo last December. Neural Information Processing At Systems (NIPS), a new machine learning tool called "generative adversarial networks" (GAN) became the focus of discussion.

The Generic Confrontation Network (GAN) was invented by Open AI scientist Ian Goodfellow, a system of two neural networks: one network learns from the training set to generate new data, and the other network tries to distinguish between real and false data. Combining these two networks can work together to produce very realistic synthetic data. This method can be used to generate scenes for video games, to clearly pixelate video images, or to create a more stylish computer design.

Yoshua Bengio, one of the world's most famous machine learning experts, and a PhD supervisor at Ian Goodfellow's PhD at the University of Montreal, said at the NIPS conference that this approach is particularly exciting because it provides a computer an effective way to learn from unmarked data.

Trend 3: China's artificial intelligence boom (China's AI boom)

2018 may be a year when China began to become a major player in the field of artificial intelligence. Today's Chinese technology companies no longer stop to imitate Western technology, but identify artificial intelligence and machine learning as the next key innovation area. We can clearly feel that domestic investors' enthusiasm for investment in artificial intelligence startups continues to rise. The government is also actively promoting policy support. It is expected to invest about US\$15 billion by 2018, demonstrating the vision of the government to promote the flourishing of artificial intelligence.

The Silicon Valley Artificial Intelligence Laboratory established by Baidu has been around for some time, and has been established in some areas, such as speech recognition and natural language processing, as well as advertising optimization. Baidu's strategic layout of AI currently includes three laboratories: Silicon Valley Artificial Intelligence Lab, Deep Learning Lab and Big Data Lab. The main research areas are image recognition, speech recognition, natural language processing, robotics and big data. In September 2016, Baidu released the Baidu brain, including the PaddlePaddle deep learning platform (algorithm model), AI supercomputer (underlying technology) and Big Three core technologies. In January 2018, the Baidu artificial intelligence operating system DuerOS was launched.

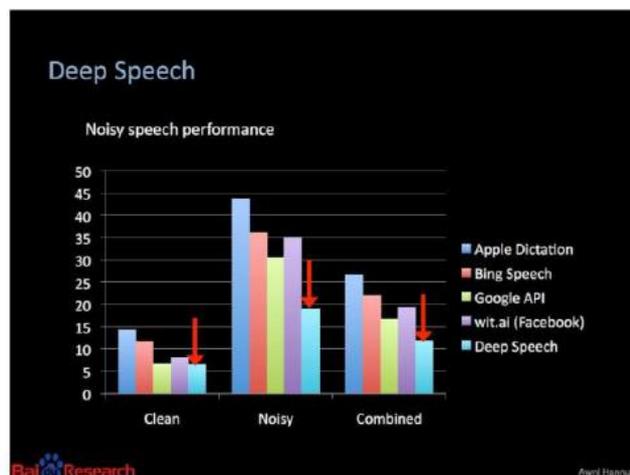


Figure 11: Comparison of speech recognition error rates between Deep Speech and Apple Dictation, Microsoft Bing Speech, Facebook wit.ai, Google API

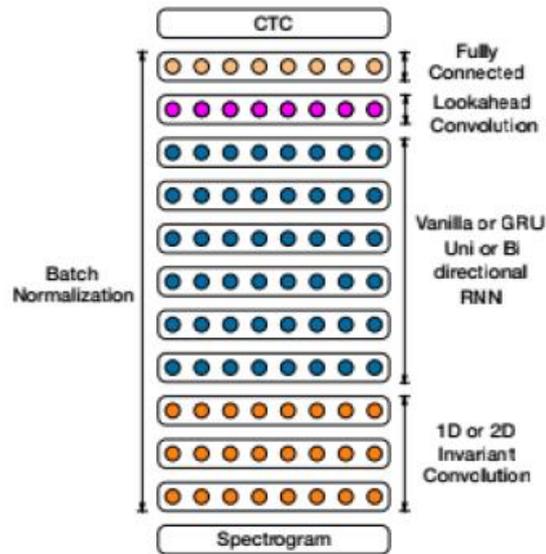


Figure 12: Deep Speech 2 for deep RNN structure in English and Chinese

In November 2015, Baidu Silicon Valley Lab developed a new generation of deep speech recognition system Deep Speech 2 under the guidance of Professor Andrew Ng. The system also uses the loss function of the Connectionist Temporal Classification (CTC) to train the network end-to-end on 16 GPUs so that the sequence of characters from the input audio can be directly predicted. Baidu researchers say that Deep Speech 2 is more prominent in noisy environments (such as in cars and in crowds). In a noisy environment, tests show that the error rate of the Baidu Deep Speech system is about 10% lower than Google's API, Facebook's Wit.AI, Microsoft's Bing Speech, and Apple's Dictation.

Trend 4: Language Learning

If you ask AI researchers what the next major goal is, they are likely to answer: language. Because it relies on technological advances in areas such as speech and image recognition, it can help computers analyze and generate languages more efficiently.

This is a long-term goal of artificial intelligence, and the prospect of computers communicating and interacting with us through language is desirable.

A better understanding of the contextual meaning of the language will give the AI system a full range of practical improvements.

AlphaGo took 37 steps in the second set of the century war with Li Shishi, and the machine chose an unusual point. DeepMind behind AlphaGo said that they only saw AlphaGo's real-time winning rate predictions. They took a few days to analyze the system carefully before they understood the meaning of AlphaGo's unusual choice at the time.

Therefore, letting artificial intelligence understand the human language and explain its decision-making behavior will in turn give scientists more inspiration. DeepMind is considering opening up AlphaGo's decision-making system to find commercially viable projects, including improvements to the Assistant and tools for medical diagnosis. And if the system can use human language, it can explain the basis of the decisions they make to the medical staff, which is more important in the medical diagnosis process.

But given the complexity of language, subtlety and multilingual ambiguity, the challenges we face are still daunting. After all, in-depth and meaningful conversations between users and smartphones are still not feasible in a short time. However, we believe that speech recognition and speech interfaces are relatively mature in terms of technology and application scenarios. And some impressive advances are going on, such as Google Assistant and Amazon Alexa. In 2018 we can look forward to seeing further developments in this area.

Trend 5: Against Artificial Hype (Backlash to the hype)

In addition to the real progress and exciting new applications, we also found that in 2016, the market's hype for artificial intelligence reached an astounding degree. While many people have confidence in the potential value of artificial intelligence technology currently being developed, reports of exaggeration and inaccuracy about artificial intelligence are equally overwhelming. We believe that artificial intelligence is now spring, and such speculation is likely to have a negative effect on the emerging industry. This situation makes the AI industry insiders feel uneasy.

The NIPS conference held a press conference for a fake artificial intelligence company called Rocket AI, which was actually a satire on the growing fanaticism and exaggeration surrounding artificial intelligence research. Professor Li Feifei said that there is a certain bubble in artificial intelligence, and it is concerned as a new thing, but it is also magnified and misinterpreted. But in fact, it has just started, and there are still room for improvement in many aspects, so it is time to prevent bubbles and speculation.

In the constant exposure of the media, large companies and investors, artificial intelligence has become the protagonist of the technology spotlight. According to EY's report, there were 33 mergers and acquisitions related to AI in 2015; 46 mergers and acquisitions related to AI in 2016. However, overexposure is also a double-edged sword. The researchers' attention will shift from the theory and research of artificial intelligence to entrepreneurship too early. At present, many startups continue to emphasize the technology of machine learning. In fact, most of them are not true. Rocket AI is one of them. In addition, companies have to face excessive premiums and valuations in mergers and acquisitions. But in the final analysis, the real problem is that when a major breakthrough does not occur, the speculators are inevitably disappointed, which leads to the startups failing to overstep due to overvaluation and the exhaustion of investment. In 2018, we should calmly look at the next development of the AI industry, and listen to some of the opposition to artificial intelligence speculation, perhaps not a bad thing. Is 2018 the best time for AI, or the worst? Without a bit of cold, you can get a plum blossom.

7. CONCLUSION

At present, the era of artificial intelligence has been centered on people's lives, and the social structure of human beings is bound to undergo major changes. The arrival of the AI era will be a new opportunity and challenge for individuals, businesses and countries. In the era of rapid development of artificial intelligence, many jobs will be replaced, so the improvement of

education and personal qualities should be changed to meet the requirements of future development. In the future, people will be liberated from bondage because of the advent of the artificial intelligence era, and truly invest in areas suitable for their own development.

REFERENCES

- [1] Xu Dan. AlphaGo artificial intelligence is coming [J]. Science 24 hours, 2016 (4): 25-27.
- [2] Zhai Xiaodong. Four directions of intelligent application in the financial field [J]. New Economic Guide, 2017 (7).
- [3] Zhang Jing. Analysis of the development status and development prospects of artificial intelligence [J]. Wireless Internet Technology, 2017(14): 134-135.
- [4] Yin Yueming, Fan Xiaomeng, Xing Beibei. Overview and Prospect of Artificial Intelligence [J]. Science and Technology and Innovation, 2017(20): 52-53.
- [5] Han Wei. Overview of the development and application of artificial intelligence technology [J]. Electronic Production, 2016(12): 95-95.