Cognitive Machine Learning

Zhongzhi Shi

Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China

Email: shizz@ict.ac.cn

Abstract

Cognitive machine learning refers to the combination of machine learning and brain cognitive mechanism, specifically, combining machine learning with mind model CAM. Three research directions are proposed in this paper, that is, emergency of learning, complementary learning system and evolution of learning.

Keywords

Cognitive Machine Learning, Emergency of Learning, Complementary Learning System, Evolution of Learning, Intelligence Science, Mind Model CAM

1. Introduction

On July 1, 2005, in commemoration of the 125th anniversary of Science, scientists summed up 125 questions, of which 94 was "What are the limitations of learning through machines?", which was interpreted as "Computers can beat the best chess players in the world, and they can grab rich information on the Internet. But abstract reasoning still goes beyond any machine".

Learning ability is the basic characteristic of human intelligence. From birth, people have been learning from objective environment and their own experience. Human cognitive ability and wisdom ability are gradually formed, developed and perfected in lifelong learning.

In 1983, Simon gave a better definition of learning: A certain long-term change that the system produces in order to adapt to the environment, makes the system can finish the same or similar work next time more effectively. Learning is the change taking place in a system; it can be either the permanent improvement of the systematic work or the permanent change in the behavior of organism. On December 12, 2015, Science magazine published a paper to show human-level concept learning through probabilistic program induction [1]. In a complicated

system, the change of learning is due to many aspects of reasons; that is to say, there are many forms of learning process in the same system.

AlphaGo is a computer program developed by Google DeepMind in London to play the board game Go. In October 2015, it had beaten a professional named Fan Hui, the European champion. In March 2016, it had beaten Lee Sedol, who is the strongest Go player in the world, in a five-game match for 4 to 1. AlphaGo's victories are a major milestone in artificial intelligence research. AlphaGo's algorithm uses a Monte Carlo tree search to find its moves based on knowledge previously "learned" by machine learning, specifically by a deep neural network and reinforcement learning [2].

Learning theory is about learning the essence of the learning process, learning the rules and constraints to study the various conditions to explore the theory and explanation. Learning is a kind of process, where individuals can produce lasting changes in the behavior by training. There are various learning theories. For over 100 years, the psychologists have provided all kinds of learning theoretical theory schools due to the difference between their own philosophical foundation, theory background, research means. These theory schools mainly include behavioral school, cognitive school and humanism school [3].

In recent years, artificial intelligence has made great progress, mainly focusing on statistics and big data. In order to solve the problem that machines have the ability of abstract reasoning and computers can learn evolution, cognitive machine learning is presented in this paper.

2. What Is Cognitive Machine Learning

Cognitive machine learning refers to the combination of machine learning and brain cognitive mechanism, specifically, combining the achievements of machine learning we have studied for many years with the mind model CAM [3]. **Figure 1** shows the cognitive machine learning.

Cognitive machine learning mainly studies the following three aspects:

1) The emergency of learning: In the process of human cognition the first step is to begin to contact with the outside world, which belongs to the stage of

Figure 1. Cognitive machine learning.

perception. The second step is to sort out and transform the materials of comprehensive perception, which belongs to the stage of concept, judgment and reasoning. We raise perceptual knowledge through visual, auditory and tactile senses to rational knowledge. After acquiring a lot of perceptual knowledge, a new concept has been formed in the human brain, which is the emergence of learning.

2) Complementary learning system: how to construct the complementary learning system between short-term memory and semantic memory?

3) Evolution of Learning: As we all know, after hundreds of thousands of years of evolution, human brain capacity is also changing. Language plays an important role in it. So learning evolution is not only to adapt to changes in the outside world, but also to change its own structure. We think it is the most important in the world to change its-owns structure.

3. The Emergency of Learning

The emergency of learning is to rise from perceptual awareness to rational knowledge, that is, conceptual learning, is a learning method as well as a form of critical thinking in which individuals master the ability to categorize and organize data by creating mind logic-based structures. This process requires both knowledge construction and acquisition because individuals first identify key attributes that would make certain subjects fall in the same category or concept. Knowledge construction is a constructive learning process in which individuals use what is familiar or what they have experienced to understand another subject matter, while knowledge acquisition is a learning process wherein a person acquires knowledge from an acknowledged expert.

The conceptual learning can be divided into two types: first-order concept generation, which is based on the similarity-recognition process, and high-order concept generation, which is based on the dissimilarity-recognition process. The first-order concept generation is related to the problem driven phase and high-order concept generation is related to the inner sense driven phase.

So far a lot of learning methods and algorithms have been proposed. Various learning methods from statistics, neural networks, fuzzy logic and deep learning can be applied to conceptual learning and pattern recognition. Here convolutional generative stochastic model (CGSM) is used as an example to illustrate the principle of emergency of learning.

Convolutional neural networks were originally based on neurocognitive machines introduced by Fukushima *et al.* into the computer field [4], later by Yann Lecun [5], has been improved and successfully applied to image detection and segmentation, object recognition and other fields.

Generally, a convolution neural network (CNN) consists of one or more convolution layers and a fully connected layer at the top (corresponding to the classical neural network), also includes non-linear mappings and some local or global pooling layers. The convolutional neural network can utilize the two-dimensional

structure of input data. Compared with other deep feed forward neural networks, convolutional neural networks need fewer parameters to estimate. It has become an attractive structure of in deep learning.CNN mainly consists of the feature extraction and the classifier. The feature extraction contains the multiple convolutional layers and sub-sampling layers. The classifier is consisted of one layer or two layers of fully connected neural networks.

However, convolutional networks deal with data noise (such as local loss, modulus), it shows a weaker. Generating random networks has the strong robustness characteristics of data noise (such as local loss, blurring, distortion, etc.) [6]. Because the generated random network has strong noise robustness and can adopt flexible frame and noise form, while the convolution network has the advantages of multi-layer invariance and spatial local correlation when extracting image visual features, and conforms to the principle of biological visual perception channel. We consider introducing convolutional network model into generating random network, and propose a convolutional generative stochastic network model (CGSM) shown in Figure 2.

Figure 2(a) describes a typical convolution generation stochastic network model architecture, which consists of three feature layers. There is no detailed description of the convolution and pooling process in each feature layer. A hidden layer may contain convolution and pooling sub-layers (e.g. h_1 and h_2), only h_3 contains convolution sub-layers. Figure 2(b) shows the computational graph. It is clear that the output of each layer is injected with no more than 50% random noise or Gaussian noise (lightning symbol) through $C\left(\tilde{X} \mid X\right)$. Then, the local supervised learning method is used to train a function to reconstruct X only from the degenerated sample $\tilde{X}$ with the highest accuracy. In this way, not only can noise data be fully learned, but also the complex task of directly modeling data to generate distribution $P(X)$ can be transformed into a more operational learning reconstruction distribution $P\left(X \mid \tilde{X}\right)$. The random elements in each layer are back propagation, where w_i' denotes the transformation of w_i, and each $X_i\,(i > 0)$ is sampled from the second reconstructed distribution $P_{\theta_2}\left(X_i \mid H_i\right)$ and the Log likelihood sum of each reconstructed distribution is used as the training objective function of the network.
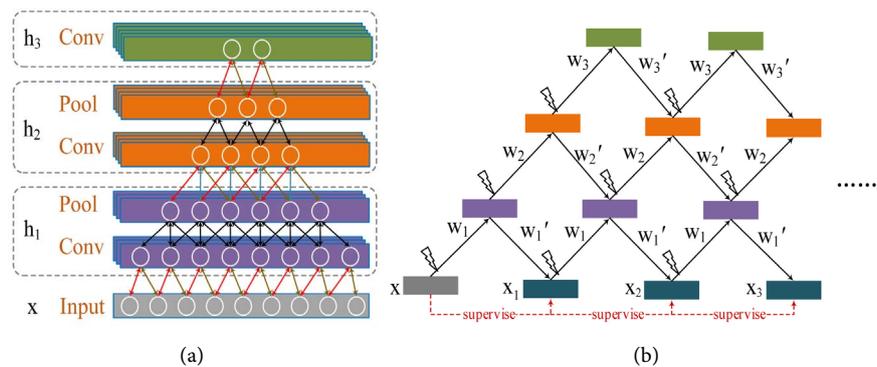

Figure 2. A typical CGSM architecture and computational graph.

In the convolution generated random model, input x is a three-dimensional array composed of n_1 two-dimensional feature maps with size $n_2 \times n_3$ and output y is a three-dimensional array composed of m_1 two-dimensional feature maps with size $m_2 \times m_3$. The convolution layer consists of a trainable filter with K size $l_1 \times l_2$. If the two-dimensional convolution is performed in Valid mode considering the boundary effect, then $m_2 = n_2 - l_1 + 1$ and $m_3 = n_3 - l_2 + 1$. In convolution layer, the output y_i is calculated by the following calculation method based on the input characteristic x_i:

$$y_{i,k} = \sigma\left(\tilde{x}_i * w_{i,k} + b_{i,k}\right) \tag{1}$$

Here, symbols $*$ represent two-dimensional convolution operations. Then x_i's reconstructed output of x_i is calculated as follows:

$$x_i' = \sigma\left(\sum_k y_{i,k} * w_{i,k}' + b_{i,k}'\right) \tag{2}$$

In the above equation, $\tilde{x}_i$ is obtained by the degeneration of x_i through $C(\tilde{x}_i \mid x_i)$, where $C(\tilde{x}_i \mid x_i)$ is injected with independent random variable $z \sim P(Z)$. In this way, x_i is converted to the operation of random variable $\tilde{x}_i$; $\sigma(\bullet)$ is an activation function, and generally $\sigma(\bullet)$ is a tanh() operated point by point. However, more complex non-linear implementations have recently been adopted, such as for natural image processing.

Convolution generates random models, usually consisting of multiple convolution generates random stacks. In order to obtain high-level feature expression, pooling technology is used after each convolution layer to reduce the size of feature graph. Common pooling technologies are Mean-Pooling and Max-Pooling. In this paper, Mean-Pooling technology is mainly used to realize Downward Pass operation. Similar to the depth-constrained Boltzmann machine network, the layer-by-layer sampling operation will also be used to learn the convolution-generated random network. The convolution-generated random network computational graph shown in **Figure 2(b)** describes the process of layer-by-layer reconstruction, including the sampling operation.

Softmax is usually used as the output of the last layer of the model when convolution generation stochastic model is used as classification and recognition task. There are obvious differences between the training process of the model and the traditional convolution network. The training process of convolution generation stochastic model is mainly divided into two stages: first, the pre-training process. The WalkBack algorithm is used to reconstruct the learning condition distribution $P(X \mid \tilde{X})$ by sampling layer by layer to approximate the actual data distribution $P(X)$ in order to obtain better robustness. Also the use of back Propagation algorithm optimizes the whole stochastic generation model globally to achieve target prediction or recognition performance.

In order to verify that convolution generated random model has more performance advantages than ordinary convolution neural network and other perception models in complex noise environment, two experiments were designed

in the experiment, namely, noise-free environment and noise environment. Noise-free environment directly uses handwritten digits in mnist0_3 data set for recognition, while noisy environment does noise processing, such as partial ambiguity, missing and so on.

From the recognition rate column in Table 1, it can be seen that the implementation forms of MLP, CNN, CGSM and SVMs can maintain a high recognition rate when recognizing a single object, while the recognition rate of CGSM is the highest. For object sequence recognition, CGSM shows a higher recognition rate than MLP, CNN and SVMs. The object recognition accuracy is 93.86%.

In the environment experiment with noise, the input data of each perception model is converted from the original input data X to the random variable $\tilde{X}$ with noise through $C(\tilde{X} \mid X)$. Noisy dataset in Table 2 is obtained by injecting 30% Bernoulli random variables and the convolution feature layer of CGSM is also injected 30% Bernoulli random variables during the experiment.

4. Complemental Learning System

In recent years, rapid progress has been made in the related fields of artificial intelligence. The benefits to developing artificial intelligence of closely examining biological intelligence are two-fold [7]. First, neuroscience provides a rich source

Table 1. Recognition rate of noise-free object sequence.

Dataset	Awareness Model	Framework	Recognition Rate (%)	
			Single	Sequence
Original Dataset	MLP	Input-H-H-H-Log	98.0272	52.66
	SVM	Linear	97.2817	29.96
		Polynomial	89.4154	0.65
		Radial Basis	98.725	56.62
		Sigmoid	98.3401	47.08
	CNN	Input-C-C-H-Log	99.7231	91.52
	CGSM	Input-Cs-Cs-H-Log	99.8024	93.8652

Table 2. Recognition rate of noisy object sequence.

Dataset	Awareness Model	Framework	Recognition Rate (%)	
			Single	Sequence
Original Dataset	MLP	Input-H-H-H-Log	96.2637	29.37
	SVM	Linear	96.3154	0.3
		Polynomial	46.3706	0
		Radial Basis	97.1561	39.84
		Sigmoid	96.6435	33.56
	CNN	Input-C-C-H-Log	99.0306	73.08
	CGSM	Input-Cs-Cs-H-Log	99.7503	92.25

of inspiration for new types of algorithms and architectures, independent of and complementary to the mathematical and logic-based methods and ideas that have largely dominated traditional approaches to AI. Second, neuroscience can provide validation of AI techniques that already exist. If a known algorithm is subsequently found to be implemented in the brain, then that is strong support for its plausibility as an integral component of an overall general intelligence system.

Artificial intelligence has been revolutionized over the past few years by dramatic advances in deep learning methods. As the field of deep learning evolved out of parallel distributed processing research into a core area within artificial intelligence, it was bolstered by new ideas, such as the development of deep belief networks, convolutional neural network.

In 1995, McClelland *et al.* proposed a complementary learning system in neocortex and hippocampus [8]. Brain effective learning requires two complementary systems: one, located in the neocortex, serves as the basis for the gradual acquisition of structured knowledge about the environment, while the other, centered on the hippocampus, allows rapid learning of the specifics of individual items and experiences. In 2016, Kumaran *et al.* extend the complementary learning system by showing that recurrent activation of hippocampal traces can support some forms of generalization and that neocortical learning can be rapid for information that is consistent with known structure [9].

Follow McClelland's idea, it is an interesting topic how to develop a complementary learning system in mind model CAM, which can implement between short-term memory and semantic memory.

According to the temporal length of memory operation, there are three types of human memory: sensory memory, short-term memory and long-term memory. The relationship among these three can be illustrated by Figure 3 [10]. First of all, the information from the environment reaches the sensory memory. If the information is attentioned then they will enter the short-term memory. It is in short-term memory, the individual to be the restructuring of the information and use and respond to. In order to analyze the information into short-term memory, you will be out in the long-term memory storage of knowledge. At the same time, short-term memory in the preservation of information, if necessary, repeat can also be deposited after long-term memory. In Figure 3, the arrows indicate the flow of information storage in three runs in the direction of the model.

In Figure 3, rehearsal refers to the psychological process in which an individual repeats the material he has previously memorized through speech in order to consolidate his memory. It is an effective method of short-term memory information storage, which can prevent short-term memory information from being disturbed by irrelevant stimuli and forgetting. After repetition, learning materials are kept in short-term memory and transferred to long-term memory, particular semantic memory.
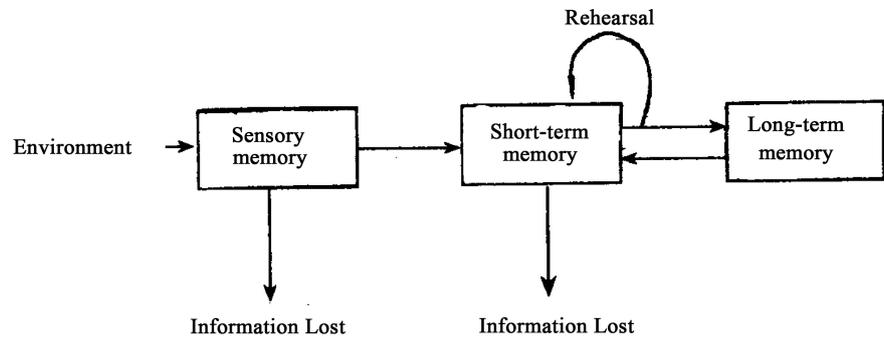
Figure 3. Human memory system.

5. Evolution of Learning

Evolution, which adapts itself to the outside world and changes its structure, is one of the most important mechanisms in the world. Evolution, learning and then advanced evolution, learning evolution, produce target, this is actually a key, random aimless machine can explore its own target through learning. Darwin founded the theory of biological evolution in the mid-19th century. Through heredity, variation and natural selection, organisms evolve and develop from low to high, from simple to complex, and from few to many.

For intelligence, the so-called evolution refers to the learning of learning, which is different from the learning of software, and its structure changes with it. This is very important, and the structural changes record the results of learning, and improve the learning method. Moreover, its storage and operation are integrated, which is difficult for computer to do so at present. The study of computer learning evolution model in this area is probably a new topic, which deserves great attention.

Studies of fossils of ancient human skulls reveal the development of the human brain, which has tripled in size over the course of two million years of evolution. With the rapid development of human intelligence, many unique cortical centers emerged in this period, such as the locomotive language center, the writing center, the auditory language center and so on. At the same time, the brain cortex has also appeared to appreciate music and painting centers, these centers have obvious positioning characteristics. Especially with the development of human Abstract thinking, the frontal lobe of human brain expands rapidly. Thus, the modern human brain is evolving continuously.

In order to make machines have human-level intelligence and break through the limitation of learning through computers, it is necessary to make machines have the function of learning evolution. Through learning, not only knowledge is increased, but also the memory structure of the machine is changed.

We consider that without evolution of learning the goal of achieving human-level general intelligence is far from completion. Here we review the principles underlying the evolution of learning, as the most fundamental to human-level machine learning.

Cognitive structure refers to the organizational form and operation mode of

cognitive activities, including a series of operational processes, such as the interaction of components and components in cognitive activities, namely the mechanism of psychological activities. Cognitive structure theory focuses on cognitive structure, emphasizing the nature of cognitive structure construction, the interaction between cognitive structure and learning [11].

Throughout the theoretical development of cognitive structure, there are Piaget's schema theory, Gestalt's insight theory, Tolman's cognitive map theory, Bruner's classification theory, Ausubel's cognitive assimilation theory and so on. Cognitive structure theory holds that the cognitive structure existing in human mind is always in the process of change and construction, and the learning process is the process of continuous change and reorganization of cognitive structure, in which the environment and individual characteristics of learners are the decisive factors. Piaget used assimilation, adaptation and balance to characterize the mechanism of cognitive structure construction. He emphasized the importance of the external environment as a whole. He believed that the rich and good multiple stimulation provided by the environment for learners was the fundamental condition for the improvement and change of cognitive structure. Modern cognitive psychologist Neisser believes that cognitive process is constructive, which includes two processes: the process of individual response to external stimuli and the process of learners' conscious control, transformation and construction of ideas and images [12]. Cognitive structure is a gradual process of self-construction under the combination of external stimulation and individual characteristics of learners.

Piaget's formalization work of intelligence development can be divided into two stages: early structuralism and later post-structuralism. The former is also called classical theory, and the latter is called the new theoretical stage. Piaget's new formal theory basically abandoned the operation structure theory and replaced it with morphism-category theory. The development series of traditional theory is from perceptual motion schema to representation schema, intuitive thinking schema to operational thinking schema. Piaget's new formal theory has become the development series of intramorphic level, intermorphic level and extramorphic level [13].

The first stage is called intramorphic level. Psychologically, it's just a simple correspondence, no combination. Common features are based on correct or incorrect observations, especially visible predictions. This is only an empirical comparison, depending on simple state transitions.

The second stage, called intermorphic level, marks the beginning of systematic combinatorial construction. Intermorphic level combination construction only occurs locally and gradually and finally does not construct a closed general system.

The last stage is extramorphic level. The main body compares morphisms by means of operation tools. Among them, the arithmetic tool is precisely to explain and summarize the content of the previous morphism.

Topos is used to describe morphism-category theory. Around 1963, Bill Law-

vere decided to figure out new foundations for mathematics, based on category theory. His idea was to figure out what was so great about sets, strictly from the category-theoretic point of view. In the spring of 1966 Lawvere encountered the work of Alexander Grothendieck, who had invented a concept of "Topos" in his work on algebraic geometry. The word "Topos" means "place" in Greek. In algebraic geometry we are often interested not just in whether or not something is true, but in where it is true [14]. A Topos is category with certain extra properties that make it a lot like the category of sets.

6. Conclusions and Future Works

Learning ability is the basic characteristic of human intelligence. This paper has proposed the cognitive machine learning to break through the limit of learning by computer. Paper pointed out the cognitive machine learning can be studied in three directions: emergency of learning, complementary learning system, learning evolution to let machine intelligence reach the human level intelligence.

Further research is on how to construct complementary learning system between short-term memory and long-term memory in CAM model. Through continuous learning, we can not only increase knowledge, but also change the structure of long-term memory. It will be of great significance to the development of artificial intelligence.

Acknowledgements

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Lake, B.M., Salakhutdinov, R. and Tenenbaum, J.B. (2015) Human-Level Concept Learning through Probabilistic Program Induction. *Science*, **350**, 1332-1338. https://doi.org/10.1126/science.aab3050

[2] Silver, D., Huang, A., Maddison, C.J., *et al.* (2016) Mastering the Game of Go with Deep Neural Networks and Tree Search. *Nature*, **529**, 484-489. https://doi.org/10.1038/nature16961

[3] Shi, Z.Z. (2017) Mind Computation. World Scientific, Singapore. https://doi.org/10.1142/10174

[4] Fukushima, K. (1980) Neocognitron: A Self-Organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position. *Biological Cy-*

bernetics, **36**, 193-202. https://doi.org/10.1007/BF00344251

[5] Lecun, Y., Bottou, L., Orr, G.B., *et al.* (1998) Efficient Backprop. *Neural Networks Tricks of the Trade*, **1524**, 9-50. https://doi.org/10.1007/3-540-49430-8_2

[6] Bengio, Y., Eric, T., Alain, G., *et al.* (2013) Deep Generative Stochastic Networks Trainable by Backprop. *Computer Science*, **2**, 226-234.

[7] Hassabi, D., Kumaran, D., Summerfield, C. and Botvinick, M. (2017) Neuroscience-Inspired Artificial Intelligence. *Neuron*, **95**, 245-258. https://doi.org/10.1016/j.neuron.2017.06.011

[8] McClelland, J.L., *et al.* (1995) Why There Are Complementary Learning Systems in the Hippocampus and Neocortex: Insights from the Successes and Failures of Connectionist Models of Learning and Memory. *Psychological Review*, **102**, 419-457. https://doi.org/10.1037/0033-295X.102.3.419

[9] Kumaran, D., Hassabis, D. and McClelland, J.L. (2016) What Learning Systems Do Intelligent Agents Need? Complementary Learning Systems Theory Updated. *Trends in Cognitive Sciences*, **20**, 512-534. https://doi.org/10.1016/j.tics.2016.05.004

[10] Shi, Z.Z. (2012) Intelligence Science. World Scientific, Singapore. https://doi.org/10.1142/8211

[11] Shi, Z.Z. (2019) Intelligence Science. 3rd Edition, Tsinghua University Press, Beijing. (In Chinese)

[12] Neisser, U. (1967) Cognitive Psychology. Prentice Hall, Upper Saddle River.

[13] Piaget, J., Henriques, G., Ascher, E. and Brown, T. (1992) Morphisms and Categories: Comparing and Transforming. Lawrence Erlbaum Associates, New York.

[14] Barr, M. and Wells, C. (1985) Toposes, Triples and Theories. Springer-Verlag, New York. https://doi.org/10.1007/978-1-4899-0021-0