Deep learning is a fast-developing field of machine learning which is based on a hierarchy of multi-layer neural network. It gets closer to the ultimate goal of artificial intelligence: imitating a human brain, which is good at processing complex input data, such as image recognition. Hinton *et al* [1, 2] developed an unsupervised greedy algorithm from deep belief networks, which was promising to solve complex optimization problems with deep architecture. Furthermore Lecun et al [3] proposed convolutional neural network which was the first learning algorithm with real multi-layer architecture. The convolutional neural network helps improve the performance of backpropagation by reducing training complexity. The current challenge of deep learning includes parallel training distributed over a CPU or GPU cluster for large scale datasets. The study how to utilize deep learning to improve old learning algorithms would be exciting.

**How Deep Learning can utilize the HPC (supercomputers) facilities to do scientific data analysis after simulation?**

The New York Times [4] reported that scientists from search giant Google and Stanford University had created the largest neural network by a 16,000-core cluster which was able to recognize cats on its own.

**Deep Learning and BigData:**

Deep Learning and Big Data are two major trends that will impact the future technology and industry. Both of them are considerably promising, but people in industry had a popular misconception that simple model of machine learning works more effectively along with Big Data. Therefore in many applications of Big Data, simple linear regression have been applied widely. But the progress of deep learning recently encourages to reassess these models.

Take speech recognition as an example, the sample size we face today are usually in the order of billions or even hundreds of billions. In an experiment carried by Google and other labs [5], researchers found that even training model like deep neural network could be sometimes underfitting. So complex model like deep learning is required in Big Data regime.

Beyond that, Deep Learning is the application of artificial intelligence and software programming through "neural networks" to develop machines that can do a wide variety of things including driving cars, working in factories, conversing with humans, translating speeches, analyzing images, video and data patterns, and diagnosing complex operational or procedural problems.

**Deep learning at Baidu**

The typical scale of training datasets: 1) Image Recognition: 100 million; 2) OCR: 100 millions; 3) Speech Recognition: 10 billions; 4)CTR: 100 billions. The training time is like a couple of weeks or months on GPU clusters. Furthermore, the training data is growing **10 times** larger every year. All these numbers indicate HPCs are required to carry deep learning tasks in big data. Baidu plans to use Heterogeneous Computers to replace HPC.

**Challenges:**

Table 1 is a summary of recent research [6]. Although deep learning is promising, there are still some remaining challenges. For example, most traditional machine learning algorithms were designed for data that would be completely cached into memory. In Big Data age, however, the data is either too large or exist in the form of streaming. Therefore, algorithms with large scale clusters that can learn from massive amounts of data are needed. In the paper [6], the authors summarized three categories of challenges: volume, variety, and velocity, which refers to large scale of data, different types of data, and the speed of streaming data, respectively.

In [7], Bengio claims (I directly quote here, in italic font. They need to be rephrased) that the challenges of deep learning come from four categories: *scaling computations, reducing the difficulties in optimizing parameters, designing expensive inference and sampling, and helping learn representations that better disentangle the unknown underlying factors of variation*.

In terms of scaling computations, *Part of the challenge is that the current capabilities of a single computer are not sufficient to achieve these goals, even if we assume that training complexity would scale linearly with the complexity of the task. This has for example motivated the work of the Google Brain team to parallelize training of deep nets over a very large number of nodes.*

*Another part of the challenge is that the increase in computational power has been mostly coming (and will continue to come) from parallel computing. Unfortunately, when considering very large datasets, our most efficient training algorithms for deep learning (such as variations on stochastic gradient descent or SGD) are inherently sequential (each update of the parameters requires having completed the previous update, so they cannot be trivially parallelized). Furthermore, for some tasks, the amount of available data available is becoming so large that it does not fit on a disk or even on a file server, so that it is not clear how a single CPU core could even scan all that data (which*

*seems necessary in order to learn from it and exploit all of it, if training is inherently sequential).*



[1] G. Hilton, S. Osindero, Y. Teh, A fast learning algorithm for deep belief nets. Neural Computation, 2006, 18(7) : 1527-1554.

[2] Y. Bengio, P. Lamblin, D. Popovici, et al, Greedy layer-wise training of deep networks. Proc. Of the 12th Annual Conference on Neural Information Processing System. 2006:153-160.

[3] Y. Lecun, L. Bottou, Y. Bengio, et al, Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998, 86(11):2278-2324.

[4] Markoff, J. (2012). How many computers to identify a cat? 16,000. *New York Times*.

[5] Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A. R., Jaitly, N., ... & Kingsbury, B. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *Signal Processing Magazine, IEEE*, *29*(6), 82-97.

[6] Chen, X. W., & Lin, X. Big Data Deep Learning: Challenges and Perspectives.

[7] Bengio, Y. (2013). Deep learning of representations: Looking forward. In*Statistical Language and Speech Processing* (pp. 1-37). Springer Berlin Heidelberg.