

# Feedforward Neural Network 01 Background

Chen Gong

10 November 2019

本节的主要目的是从一个较高的角度来介绍一下，什么是深度学习，并且给深度学习一个较好的总结，给大家一个较好的印象。机器学习是目前最火热的一个研究方向，而机器学习大致可以分为，频率派和贝叶斯派。频率派逐渐演变出了统计机器学习，而贝叶斯派逐渐演变出了 PGM，也就是概率图模型。下面我们分开进行描述。

## 1 频率派

统计机器学习方法基本就是由频率派的估计思想得到的。统计机器学习方法大概可以分成四种。

1. 正则化： $L_1, L_2$  也就是之前提到的 Lasso 和岭回归，这实际上并没有产生新的模型，而是在之前模型的基础上进行了改进。我们可以把它描述为 Loss function + regularized。用来抑制训练的过拟合。

2. 核化：最著名的就是我们之前提到的，Kernel Support Vector Machine (SVM) 了。

3. 集成化：也就是 Adaboost 和 Random Forest。

4. 层次化：层次化主要就是我们指的 Neural Network，也就是神经网络，神经网络进一步发展就得到了我们现在研究的深度学习。而神经网络中比较著名的几类就是：1. 多层感知机 (Multiple Layer Perception); 2. Autoencoder; 3. CNN; 4. RNN。这几个组合起来就是我们经常听到的 Deep Network。

## 2 贝叶斯派

贝叶斯派的估计方法就演化得到了概率图模型 (Probability Graphic Model)。他们大致可以分成以下三类：

1. 有向图：Bayesian Network，深度增加之后就是 Deep Directed Network，包括大家听得很多的：Variational Automation Encode (VAE)，Generative Adversarial Network (GAN) 和 Sigmoid Belief Network 等等。

2. 无向图：Markov Network，深度增加之后就是 Deep Boltzmann Modeling，这就是我们的第二类图模型。

3. 有向图和无向图混合在一起，就是我们常说的 Mixed Network，主要包括，Deep Belief Network 等等。

而上述几个图模型，加上深度之后就是我们常说的 Deep Generative Network，深度生成模型。

在我们狭义的深度学习的理解中，什么是深度学习，实际上就是统计学习方法中的层次化中的 Deep Network。而广义的深度学习，还应该包括，Deep Generative Network。而实际上绝大多数

的深度学习者都不太了解 Deep Generative Network，确实涉及到贝叶斯的理论，深度学习就会变得很难。而且它的训练也会变得非常的复杂。

# Feedforward Neural Network 02 Development

Chen Gong

11 November 2019

本节主要是来讨论一下，机器学习的发展历史，看看如何从感知机到深度学习。

## 1 从时间的发展角度来看

1958 年：up，首次提出了 Perceptron Linear Algorithm (PLA)，这里就是我们机器学习的开端了。

1969 年：down，Marvin Lee Minsky 提出了，PLA has a limitation。因为 PLA 算法解决不了 non-linear 问题，比如说 XOR 问题。非常戏剧的是，这一年，Marvin Lee Minsky 获得了图灵奖，他也是“人工智能之父”，第一位因为 AI 而获得图灵奖的科学家。

1981 年：up，学者提出了 Multiple Layer Perceptron (MLP)，可以用来解决非线性的问题，就是最初的 Feedforward Neural Network。

1986 年：up，Hinton 提出了将 Back Propagation (BP) 算法和 MLP 完美的融合在了一起，并且发展出了 Recurrent Neural Network (RNN) 算法。

1989 年：up，提出了 CNN。但是也迎来了人工智能的寒冬。down，在这一年中提出了一个 Universal Approximation theorem，也就是一个大于 1 层的 Hidden Layer 就可以用来拟合任何的连续函数。那么这是就提出了一个疑问：1 layer is OK, why deep? 并且，在 BP 算法中，随着深度的增加还会出现梯度消失的问题。

1993 年和 1995 年，down，这一年中 Support Vector Machine (SVM) + Kernel + Theory，获得了很好的效果。并且，Adaboost 和 Random Forest 等 Ensemble algorithm 流派的提出，获得了很好的效果。

1997 年，up，提出了 LSTM，但是远不足以止住深度学习发展的颓势。

2006 年，up，Hinton，提出了 Deep Belief Network (RBM) 和 Deep Auto-encoder。

2009 年，up，GPU 的飞速发展。

2011 年，up，Deep Learning 运用到了语音 (Speech) 中。

2012 年，up，斯坦福大学李飞飞教授，开办了一个非常重要的比赛和数据库 ImageNet。

2013 年，up，提出了 Variational Autoencoder (VAE) 算法。

2015 年，up，提出了非常重要的 Generative Adversarial Network (GAN)。

2016 年，up，围棋上 AlphaGo 彻底引爆了 Deep Learning。

2018 年，up，提出了重要的 Graph Convolutional Network (GCN)，传统的神经网络是连接主义的，而 GCN 中将符号主义和连接主义进行了联合，使之具有推理的功能。

## 2 总结

其实 Deep Learning 的崛起是很多因素融合的结果。这些年来，主要都是在实践上的发展，而在机器学习理论上基本没有什么进步。它的发展得益于以下几点：1. data 的则增加，big data 时代的到来；2. 分布式计算的发展；3. 硬件水平的发展。其实最主要的说白了就是效果，效果比 SVM 要更好，就占据了主要的地位。

计算机学科就是一门实践为主的科学，现在在实际上取得了很好的效果。之后随着理论研究的不断深入，我们一定可以不断的完善理论知识。之后 AI 方向的研究，也将是以深度学习为主流，而其他机器学习学派的知识 and 优点将不断地丰富深度学习，扩充深度学习，来给它更强大的效果。

# Feedforward Neural Network 03 Non-Linear Problem

Chen Gong

12 November 2019

实际上在 1958 年就已经成功的提出了 Perceptron Linear Analysis (PLA)，标志着人工智能的正式诞生。但是，Minsky 在 1969 年提出 PLA 无法解决非线性分类问题，让人工智能陷入了 10 年的低谷。后来的发展，人们开始寻找到越来越多的，解决非线性分类问题的方法。于是，我们提出了三种解决非线性问题的方法。

## 1 Non-Transformation

这实际上就是一种明转换，将向量从 input space 转换到 feature space，可以写做  $\phi = \mathcal{X} \mapsto \mathcal{Z}$ 。在 Conver's theory 中提出，高维空间比低维空间更加容易线性可分。很显然对于一个异或问题 (XOR) 来说，我们将  $x = (x_1, x_2) \xrightarrow{\phi} z = (x_1, x_2, (x_1 - x_2)^2)$

$$\begin{array}{ccc|ccc} 0 & 1 & \longrightarrow & 1 & 0 & 1 & 1 & \longrightarrow & 1 \\ 1 & 0 & \longrightarrow & 1 & 1 & 0 & 1 & \longrightarrow & 1 \\ 1 & 1 & \longrightarrow & 0 & 1 & 1 & 0 & \longrightarrow & 0 \\ 0 & 0 & \longrightarrow & 0 & 0 & 0 & 0 & \longrightarrow & 0 \end{array} \xrightarrow{\phi} \quad (1)$$

很显然在三维空间中，进行空间映射后，就会变得比较容易进行线性划分了。可以自己画图来进行验证，这里不再作图。

## 2 Kernel Method

这实际上是一种暗转的思路，也就是令  $K(x, x') = \langle \phi(x), \phi(x') \rangle$ ，在这个核函数中实际上隐藏了一个  $\phi$ ，而  $x, x' \in \mathcal{X}$ 。

## 3 Neural Network

神经网络算法实际上就是一个 Multit-Layer Perceptron (MLP)，有时也会被称为，Feedforward Neural Network (FNN)，所以大家在其他书上见到这几种描述都不要感到意外。我们以 XOR (位运算) 为例吧。在我们的逻辑运算中，大致有四种运算方法。

$$\begin{array}{cccc} XOR & OR & AND & NOT \\ \oplus & \vee & \wedge & \neg \end{array} \quad (2)$$

而后三种运算为基础运算，因为异或运算实际上是可以由后三种运算组成，也就是  $x_1 \oplus x_2 = (\neg x_1 \wedge x_2) \vee (x_1 \wedge \neg x_2)$ 。实际上就是先做两个与运算，然后做一个或运算。把一个线性不可分的东西来分层实现，将特征空间进行了分解而已。然后，在分层运算中插入了激活函数，来达到非线性映射的效果。这部分内容，比较的简单，而且网上也有大量的资料，此处就不再做过多的阐述。

实际上神经网络就是一个有向无环图。所以，某种意义上说可以引入概率图的模型，当然这就是后话了。