

# 第一章 绪论

## 1.1 统计学习

### 1.1.1 统计学习的定义

#### 1. 学习

学习是人类具有的一种普遍智能行为，是贯穿于每个人一生的可持续动态过程。但究竟什么是学习，长期以来却众说纷纭，社会学家、逻辑学家和心理学家都各有其不同的看法，并未形成统一的观点。本书采用百度百科给出的定义：狭义来讲，学习是指“通过阅读、听讲、研究、观察、理解、探索、实验、实践等手段获得知识或技能的过程，是一种使个体可以得到持续变化（知识和技能、方法与过程、情感与价值的改善和升华）的行为方式”；广义来讲，学习是指“人在生活过程中，通过获得经验而产生的行为或行为潜能的相对持久的行为方式”。

尽管学习的狭义定义和广义定义有所不同，但二者之间具有共同的部分，即学习包括知识的学习、行为的学习和技能的学习，以及抽象逻辑的理解和空间思维的想象。这就是学习的本质。

#### 2. 机器学习

让机器像人类那样具有智能、思维和学习能力一直是科学工作者为之奋斗的目标。现代计算机科学之父英国人阿兰·图灵于1950年发表了一篇题为《机器能思考吗？》的论文，在这篇论文里，图灵第一次提出“机器思维”的概念，以及测试机器是否具有思维的实验，即一个人在不接触对方的情况下，通过一种特殊的方式，和对方进行一系列的问答，如果在相当长的时间内，他无法根据这些问题判断对方是人还是计算机，那么，就可以认为这个计算机具有同人相当的智力，即这台计算机是能思维的。这就是著名的“图灵测试”（Turing Testing）。

机器能否像人类一样能具有学习能力？1959年，美国的塞缪尔（Samuel）设计了一个下棋程序，它可以在不断的对弈中改善自己的棋艺。4年后，这个程序战胜了设计者本人。又过了3年，这个程序战胜了美国一个保持8年之久的常胜不败的冠军。这个程序向人们展示了机器也可以像人类一样具有学习的能力。

为了便于科学研究，有必要先对机器学习给出定义。但究竟什么是机器学习，长期以来也是众说纷纭。

最直观的定义，机器学习是研究如何使用机器来模拟人类学习活动的一门学科。

Langley（1996）定义的机器学习是“机器学习是一门人工智能的科学，其主要研究对象是人工智能，特别是利用经验提高算法性能”（Machine learning is a science of the artificial. The field's main objects of study are artificial, specifically algorithms that improve their performance

with experience)。

Tom Mitchell (1997) 定义的机器学习是“机器学习是借助于经验自动改进计算机算法的研究”(Machine Learning is the study of computer algorithms that improve automatically through experience)。

Alpaydin (2004) 对机器学习的定义是“机器学习是计算机利用示例数据或以往的经验优化的性能标准。”(Machine learning is programming computers to optimize a performance criterion using example data or past experience)。

科普中国百科科学词条编写与应用工作项目定义机器学习是一门多领域交叉学科, 涉及概率论、统计学、逼近论、凸分析、算法复杂度理论等多门学科, 专门研究计算机怎样模拟或实现人类的学习行为, 以获取新的知识或技能, 重新组织已有的知识结构使之不断改善自身的性能。

计算机科学, 特别是人工智能的研究者一般公认 Simon 对学习的论述: “如果一个系统能够通过执行某个过程改进它的性能, 这就是学习。”这是一个相当广泛的说明, 其要点是“系统”, 它涵盖了计算系统、控制系统以及人系统等。

### 3. 统计学习

根据百度百科的划分, 机器学习是人工智能研究的分支之一, 它的发展过程大体上经历了 4 个阶段:

第一阶段: 20 世纪 50 年代中叶~60 年代中叶, 这一时期属于机器学习的热烈时期;

第二阶段: 20 世纪 60 年代中叶~70 年代中叶, 这一时期属于机器学习的冷静时期;

第三阶段: 20 世纪 70 年代中叶~80 年代中叶, 这一时期属于机器学习的复兴时期;

第三阶段: 这一时期属于机器学习的最新阶段, 始于 1986 年, 其标志为 Vapnik 的统计学习理论的本质(The nature of statistical learning theory, 也称统计机器学习, 即 statistical machine learning)。从此, 特别是近几年, 统计机器学习是被广泛应用的机器学习方法, 是机器学习比较热门的一个研究方向。但什么是统计机器学习呢? 我们综合本书列出的文献[1][2][3][4]以及百度百科关于统计机器学习的描述, 给出如下定义:

统计学习是概率论、统计学、信息论、计算理论、最优化理论和计算机科学等多个领域的交叉学科, 并且在发展中形成了独立的理论体系和方法论; 是关于计算机基于数据构建的概率统计模型, 并运用模型对数据进行预测和分析的一门科学。

#### 1.1.2 统计机器学习概述

统计机器学习研究的问题一般是: 给定一带标签的、有限的所有观测对象的集合(也称为训练样本集合, 训练样本集合是有限的, 这是由于我们观察能力的限制, 只能获得所有样本的一个有限的样本子集), 需要从训练集合中, 基于概率论、统计学、信息论、计算理论、最优化理论和计算机科学等多个领域知识, 构建概率统计模型并运用模型对数据进行预测与分析, 使得预测结果尽可能接近它的真实标签。具体来讲, 对给定一带标签的、有限的所有观测对象的集合, 学习的目标是根据这样一个样本集合, 从一个事先给定的分类器集合中挑选出一个分类器, 使得分类器对从同一个分布中随机抽取的样本在给定的一个损失评价下的

风险最小。

通过上述对统计机器学习的概述，我们可以发现统计机器学习方法主要涉及四个问题：训练样本（或称之为数据）、模型假设、训练策略和训练算法。

### 1. 训练样本

样本是统计学习的基础，也是统计学习的对象。数据是样本的组成元素，对样本的训练，也就是对数据的分析和学习，即提取数据特征、发现数据之间的因果或关联关系，抽象出概率模型，运用模型对数据进行预测与分析。由此可以看出，数据贯穿于统计机器学习的始终。

随着互联网催生的大数据的出现，统计机器学习中的数据从数据的种类、数据描述形式、数据来源和数据内容等也都发生了巨大变化，总结如下：

#### （1）数据种类。

作为统计学习的研究对象，数据的种类是繁多的，比如各种数字、文字、符号、图片、图标、音频、视频以及它们可能的组合。

#### （2）数据来源。

在当今的信息社会，数据来源相当广泛：传感器产生的数据，科学计算中间结果和最终结果所产生的数据，互联网数据，数据库中的数据，能源、交通、银行、商业、医疗、教育和娱乐等行业产生的数据。物联网、云计算、移动互联网、车联网、手机、平板电脑、PC 以及各种各样的传感器，无一不是数据来源或者承载的方式。总而言之，有信息交互的地方，就一定会产生数据。

#### （3）数据巨量。

无所不在的移动设备、RFID、无线传感器每时每刻都在产生数据，数以亿计的用户在互联网服务时时刻刻在产生巨量数据交互。根据百度文库提供的信息<sup>[5]</sup>，截至 2012 年，数据量已经从 TB(1024GB=1TB)级别跃升到 PB(1024TB=1PB)、EB(1024PB=1EB)乃至 ZB(1024EB=1ZB)级别。而到了 2020 年，全世界所产生的数据规模将达到今天的 44 倍。每一天，全世界会上传超过 5 亿张图片，每分钟就有 20 小时时长的视频被分享。对特定的用户来讲，这些巨量的数据有多少是可用的，如何实时汲取有用的数据，也是一个具有挑战性的研究课题。

这些数据无论多么巨量、种类多么繁多，从统计机器学习的角度来讲，首先在满足同类数据应具有一定统计规律的前提外，大数据的出现，给统计机器学习带来了新的问题：

问题之一，巨量数据之中包含大量的无用信息，统计机器学习如何快速地、实时地从巨量数据之中剔除无用信息，获取有用数据，提高统计机器学习的效率；

问题之二，传统的统计机器学习方法，如何更新，以便应对大数据的挑战；

问题之三，从传统的发现数据中的因果知识发展到发现数据中的因果知识和数据之间的关联关系。

### 2. 模型假设

#### （1）输入、输出空间。

在统计机器学习中，我们将学习系统的输入与输出所有可能取值的集合称为输入空间（Input Space）和输出空间（Output Space）。输入、输出空间可以是同一个空间，也可以是不同的空间；可以是有限元素的集合，也可以是无限元素的集合。

## (2) 假设空间。

假设空间也称作模型空间，即从输入空间到输出空间的所有映射函数的集合。这些函数对统计机器学习来讲，往往是一个后验概率或者是一个映射函数。

统计机器学习的目标是从假设空间中寻找一个最优的模型，基于这个最优的模型，如何将样本从输入空间转化到输出空间。

## 3. 训练策略

模型所在空间也就是假设空间，往往包含无穷多个满足假设的可选模型，如何从假设空间中选择一个最优模型，应该采用怎样的选择标准？这就是模型选择应该解决的问题。一般采用损失函数和风险函数来制定模型选择策略，将模型选择转化为一个最优化问题来求解。

## (1) 损失函数。

用  $X$  表示输入空间中的一个变量， $f$  表示从假设空间选取的一个模型，则  $f(X)$  表示从输入空间到输出空间的预测值， $Y$  表示输出空间中的真实值。用一个称之为损失函数的非负实值函数  $F(Y, f(X))$  来度量预测误差的程度。常用的损失函数包括：

- 0-1 损失函数：

$$F(Y, f(X)) = \begin{cases} 1 & Y \neq f(X) \\ 0 & Y = f(X) \end{cases}$$

- 平方损失函数：

$$F(Y, f(X)) = (Y - f(X))^2$$

- 绝对损失函数：

$$F(Y, f(X)) = |Y - f(X)|$$

一般情况下，损失函数越小，模型就越好。统计机器学习的目标就是，通过对样本的学习，使得损失函数极小，即： $\min_{f \in \text{假设空间}} F(Y, f(X))$ 。

## (2) 经验风险函数。

给定一个样本训练集：

$$W = \{(x_i, y_i), i = 1, 2, 3, \dots, N\}$$

式中， $x_i \in X$ ， $y_i \in Y$ ，模型  $f(X)$  关于样本训练集的平均损失称为经验风险，记为：

$$R_{\text{emp}}(f) = \frac{1}{N} \sum_{i=1}^N F(y_i, f(x_i))$$

一般情况下，风险函数越小，模型就越好。统计机器学习的目标就是，通过对样本的学习，使得风险函数极小，即： $\min_{f \in \text{假设空间}} \frac{1}{N} \sum_{i=1}^N F(y_i, f(x_i))$ 。但是，当样本容量较小时，经验风险最小化学习会产生过拟合 (overfitting)。

## (3) 结构风险函数。

由于经验风险最小化学习会产生过拟合 (overfitting)，通常我们也会在损失函数中加上正则化项，从而降低模型的复杂性，提高模型的泛化能力，拒绝 overfitting，形成结构风险最小化学习方法，即：

$$\min_{f \in \text{假设空间}} \frac{1}{N} \sum_{i=1}^N F(y_i, f(x_i)) + \lambda J(f)$$

式中， $J(f)$  表示模型复杂度， $\lambda \geq 0$  是系数，用以权衡经验风险和模型复杂度。

#### 4. 训练算法

学习算法是用来解决最优化问题的方法。在给定损失函数或风险函数后，如何快速找到损失函数约定条件下的最优解就是学习算法需要解决的问题。常用的学习算法包括梯度下降、拟牛顿法等。

### 1.1.3 统计机器学习的常用学习算法

统计机器学习包括监督学习、无监督学习、半监督学习、提升学习和强化学习，以及当前比较热门的深度学习机器学习算法。下面逐一进行简单介绍。

#### 1. 监督学习 (Supervised Learning)

监督式学习或有教师学习是机器学习中的一个方法，可以由一组已知类别的样本（训练集）中学到或建立一个模型（Learning Model），并依此模型推测新的实例。正如人们通过已知病例学习诊断技术那样，计算机要通过学习才能具有识别各种事物和现象的能力。人们根据输入、输出变量的不同类型，对预测任务给予不同的名称：输入变量和输出变量均为连续变量的预测问题称为回归问题；输出变量为有限个离散变量的预测问题称为分类问题；输入变量与输出变量均为变量序列的预测问题称为标注问题。由于本书主要研究监督学习在机器视觉中的应用，因此对于监督学习，我们主要讨论分类问题。

分类是监督学习的一个核心问题。在监督学习中，输出变量为有限个离散变量的预测问题称为分类问题，预测问题便成为分类问题。分类问题包括学习和分类两个过程。在学习过程中，根据已知的训练数据集利用有效的学习方法学习一个分类器（概率密度或决策函数）；在分类过程中，利用学习的分类器对新的输入实例进行分类。其框架结构如图 1-1 所示。

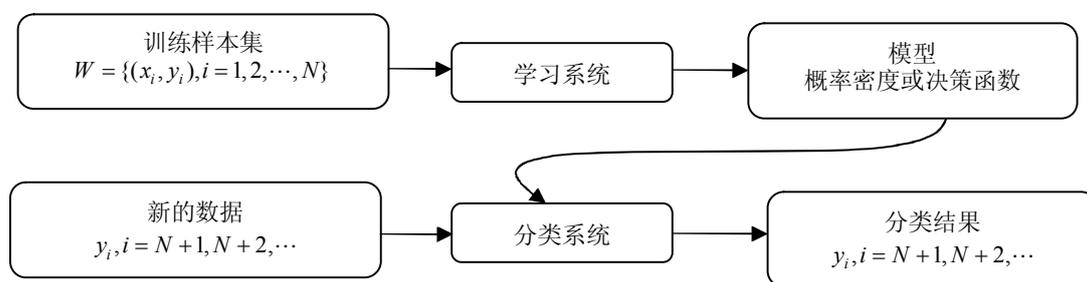


图 1-1 分类学习框图

常用的具有代表性的分类算法有：K 近邻 (K-Nearest Neighbor, KNN)、神经网络 (Neural Network)、决策树 (Decision Tree)、支持向量机 (Support Vector Machine, SVM)、提升方法 (Boosting) 和深度学习 (Deep Learning)。这些算法将在后面的章节中介绍。

## 2. 无监督学习 (Unsupervised Learning)

在监督学习中,目标是学习从输入到输出的映射关系,其中输出的正确值(即输出空间向量)已经由指导者或者系统提供。然而,无监督学习却没有这样的指导者,只有原始输入数据。我们的目标是发现输入数据中的规律,即发现输入数据的簇或分组。其特点是仅对此种网络提供输入范例,而它会自动从这些范例中找出其潜在类别规则。当学习完毕并经测试后,也可以将之应用到新的案例上。

无监督学习包括聚类(Clustering)和降维(Dimensionality Reduction)两类任务,常用的代表性的算法有:K-均值(K-means)、层次聚类(Hierarchical Clustering)、主成分分析(Principal Component Analysis, PCA)和相关分析法(Correlation Analysis)。

## 3. 半监督学习 (Semi-Supervised Learning) [6]

在许多机器学习的实际应用中,很容易找到海量的无类标签的样例,但需要使用特殊设备或经过昂贵且用时非常长的实验过程进行人工标记才能得到有类标签的样本,由此产生了极少量的有类标签的样本和过剩的无类标签的样例。因此,人们尝试将大量的无类标签的样例加入到有限的有类标签的样本中一起训练来进行学习,期望能对学习性能起到改进的作用,由此产生了半监督学习。

半监督学习的形式化描述为:给定一个来自某未知分布的样本集  $S$ , 此样本集由两部分组成, 其一是已标签样本集  $L = \{(x_i, y_i), i = 1, 2, \dots, l\}$ , 其二是未标签样本集  $U = \{x'_i, i = 1, 2, \dots, u\}$ , 则  $S = L \cup U$ , 构造  $X$ , 使得  $x_i \cup x'_i \in X$ , 其中,  $x_i$  和  $x'_i$  均为同维向量,  $y_i \in Y$  为样本  $x_i$  的标签。半监督学习就是在样本集  $S$  上寻找最优的学习器, 希望得到函数  $f: X \rightarrow Y$  可以准确地对样本  $x$  预测其标签  $y$ 。这个函数可能是参数的, 如最大似然法; 可能是非参数的, 如最邻近法、神经网络法、支持向量机等。

## 4. 提升学习 (Boosting Learning) [7]

提升学习是一种用来提高学习算法准确度的方法, 这种方法的基本思想是 Schapire 基于 PAC 的弱可学习理论证明了一个定理: 如果一个概念是弱可学习的, 充要条件是它是强可学习的。它通过构造一个预测函数(弱分类器)系列, 然后以一定的方式将它们组合成一个预测函数(强分类器), 达到把弱学习算法提升为强学习算法的目的。1989年, Schapire 提出了第一个可证明的多项式时间 Boosting 算法, 对这个问题作出了肯定的回答。一年后, Freund 设计了一个高效得多的通过重取样或过滤运作的 Boosting by Majority 算法。

AdaBoost 是 Boosting 家族中的基础算法, 其家族中的大部分扩展算法都由它得来。所以, 在这里有必要对 AdaBoost 算法做一简单介绍。

AdaBoost 是一种迭代算法, 其核心思想是针对同一个训练集训练不同的分类器, 即弱分类器, 然后把这些弱分类器集合起来, 构造一个更强的最终分类器。算法本身是改变数据分布实现的, 它根据每次训练集之中的每个样本的分类是否正确, 以及上次的总体分类的准确率, 来确定每个样本的权值。将修改权值的新数据送给下层分类器进行训练, 然后将每次训练得到的分类器融合起来, 作为最后的决策分类器。其基本结构如图 1-2 所示。

## 5. 强化学习 (Reinforcement Learning)

我们将任何独立的能够思想并可以同环境交互的实体称为智能体。所谓强化学习就是智

能体从环境到行为映射的学习，从动物学习、参数扰动自适应控制等理论发展而来，其基本原理是：如果 Agent 的某个行为策略导致环境正的奖赏（强化信号），那么 Agent 以后产生这个行为策略的趋势便会加强。Agent 的目标是在每个离散状态发现最优策略以使期望的折扣奖赏和最大。

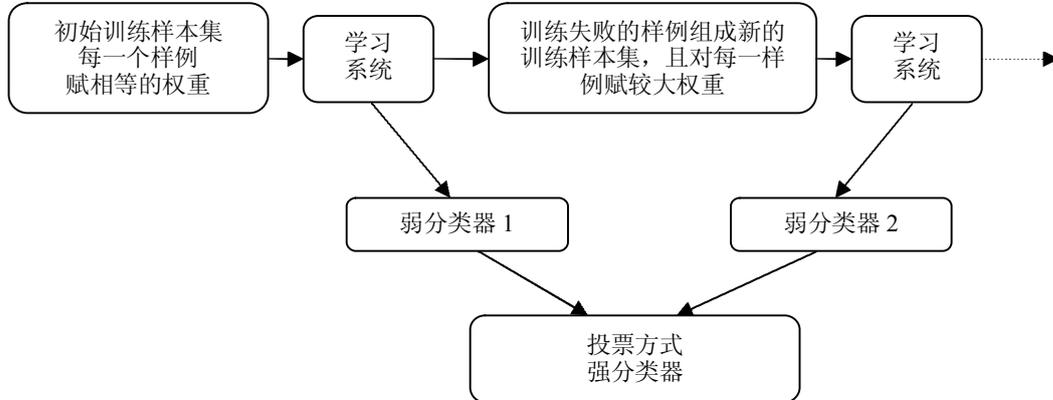


图 1-2 提升学习原理框图

强化学习把学习看作试探评价过程，Agent 选择一个动作用于环境，环境接受该动作后状态发生变化，同时环境对 Agent 所产生动作的好坏作一种评价，即强化信号（奖或惩），并反馈给 Agent，Agent 根据强化信号和环境当前的状态再选择下一个动作，选择的原则是使受到正强化（奖）的概率增大。强化学习的原理如图 1-3 所示。

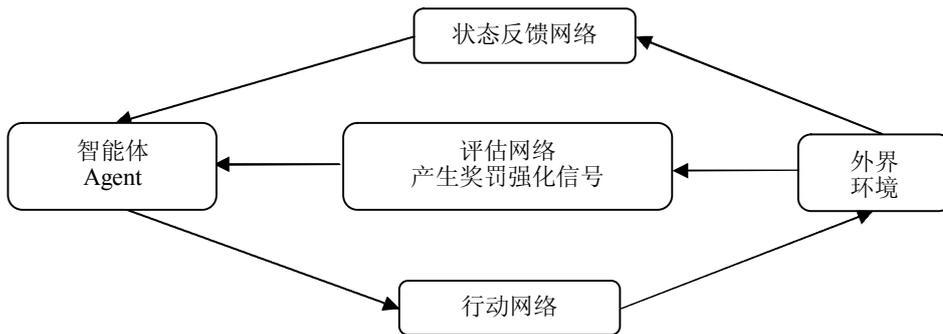


图 1-3 强化学习原理框图

## 6. 深度学习（Deep Learning）

20 世纪 80 年代末期，用于人工神经网络的反向传播算法（也叫 Back Propagation 算法或者 BP 算法）的发明，给机器学习带来了希望，掀起了基于统计模型的机器学习热潮。

随着机器学习的算法深入到结构化层次上，研究人员发现，传统的 BP 神经网络，虽然也被称作多层感知机（Multi-layer Perceptron），但实际上是一种只含有一层隐层节点的浅层模型，且传统的 BP 神经网络机器学习算法存在如下问题：

- (1) 从顶层到底层，残差信号越来越小，出现所谓的 Gradient Diffusion（梯度扩散）。

(2) 容易收敛到局部最小值。

(3) 一般情况下，只能用有标签的数据来训练网络。

2006年，Geoffrey Hinton 提出了在非监督数据上建立多层神经网络的一个深度学习方法（Wake-Sleep 算法），以弥补上述 BP 神经网络的局限性。Wake-Sleep 算法分为醒（wake）和睡（sleep）两个部分。wake 部分从底层开始，采用无标定数据（有标定数据也可），使用自下而上的非监督学习，将本层的输出作为下一层的输入，一层一层地往顶层训练，分层训练各层参数；sleep 部分自顶层开始，采用带标签的数据，使用自上而下的监督学习，对网络参数进行微调。

## 1.2 机器视觉

### 1.2.1 相关概念

#### 1. 视觉

基于百度百科给出的描述：人类是通过感官来认知世界的。一般情况下，人们通过视觉、听觉、嗅觉、触觉、味觉这五种感受而形成对信息的感知。在人类认知客观世界的五大感官之中，视觉是人们认知接受信息的主要渠道。经调查，在这五种感官中，通过视觉来认知信息占 70%，听觉占 20% 左右，其余感官的获取量仅占 10%。

人类的视觉系统包括眼球和大脑皮层枕叶，以及两者之间的视路系统。当人们看东西时，物体的影像经过瞳孔和晶状体，落在视网膜上，视网膜上的视神经细胞在受到光刺激后，将光信号转变成生物电信号，通过神经系统传至大脑，再根据人的经验、记忆、分析、判断、识别等极为复杂的过程而构成视觉，在大脑中形成物体的形状、颜色等概念。

视觉系统是人们接触外界信息最常用的器官，它能将现象作理性的分析、联想、诠释和领悟，并且在视觉过程中对信息轮廓进行辨析、判断。视觉不仅是心理与生理的知觉，更是创造力的根源，其经验来自对四周环境的领悟与辨析。

#### 2. 计算机视觉

根据“科普中国”百科科学词条编写与应用工作项目的审核，计算机视觉是一门研究如何使机器“看”的科学，更进一步地说，就是指用摄影机和电脑代替人眼对目标进行识别、跟踪和测量等机器视觉，并进一步做图形处理，使其成为更适合人眼观察或传送给仪器检测的图像。从本质上来讲，计算机视觉是使用计算机及相关设备对生物视觉的一种模拟。它的主要任务就是通过对采集的图片或视频进行处理、分析和识别，以获得相应场景的三维信息。它的发展经历了如下几个阶段<sup>[8][9]</sup>：

(1) 计算机视觉始于 20 世纪 50 年代的统计模式识别，其目的是分析和识别二维图像目标。此时，其应用领域并不广泛，主要集中于光学字符识别、远程照片等平坦的制造物表面、景深较小的显微图片对比度较低的、航空图片的分析和解释等。

(2) 20 世纪 60 年代，Roberts 通过计算机程序由数字图像中推导出诸如立方体、楔形体、棱柱体等多面体的三维结构和空间安排，并对物体形状及物体的空间关系进行描述。比较有代

表性的是 Guzman 设计的 SEE 系统，该系统用边缘所构成的角特征作为主要信息源，将线段的各种连接情况进行分类。归纳出一系列启发式规则，并依靠对交角类型分析将相邻的区域联系在一起，构成一个完整的目标。

(3) 20 世纪 70 年代初计算机视觉的研究领域普遍出现低谷。为了摆脱这一困境，计算机视觉迫切地需要有一个统一的理论框架作指导。70 年代中期到 80 年代初期，计算机视觉的研究领域首次出现了一个 Marr 视觉计算理论框架，将视觉系统从概念上分成几个独立的模块，各自完成不同的计算任务。这一视觉计算理论框架对计算机视觉的发展起到了积极的推动作用。

(4) 20 世纪 80 年代中期以后，计算机视觉获得蓬勃的发展，新的概念、方法、理论不断涌现，而且计算机视觉的研究经历了从实验室走向实际应用的发展阶段。最有代表性的作品是英属哥伦比亚大学教授大卫·罗伊 (David Lowe) 提出 SIFT (Scale Invariant Feature Transform, 尺度不变特征变换) 的算法，表明计算机视觉研究局部特征描述符的崛起。

(5) 2000—2015 年是面向描述符的机器学习时代。研究人员逐步从视觉几何恢复问题转向更为雄心勃勃的目标识别问题，涌现了词袋 (Bag of Words, BoW)、空间金字塔 (Spatial Pyramids)、矢量量化 (Vector Quantization)、方向梯度直方图 (HOG)、DPM 以及卷积神经网络和深度学习，以及在计算机视觉的各个阶段使用的各种机器学习工具。

### 3. 机器视觉

根据百度百科的定义：机器视觉就是用机器代替人眼来做测量和判断。机器视觉系统是指通过机器视觉产品 (即图像摄取装置) 将被摄取目标转换成图像信号，传送给专用的图像处理系统；图像系统对这些信号进行各种运算来抽取目标的特征，进而根据判别的结果来控制现场的设备动作。另外，有的学者也认为机器视觉是计算机视觉在工厂自动化的一个应用。

## 1.2.2 机器视觉概述

### 1. 机器视觉研究内容

一般来讲，机器视觉系统包括：光路系统、图像传感器、图像彩卡、图像专门处理器、主控计算机、监视器、输入输出单元、执行机构等。其主要任务就是获取场景图像，对场景图像进行分割、识别、理解，作出相应的决策。由此可见，机器视觉包括如下主要研究内容：

#### (1) 图像获取。

图像获取通常由图像传感器 (如：照相机、摄像机等)、图像采集卡以及计算机系统等组成。主要任务就是获取场景信息，并传送给计算机，其作用相当于人的眼睛。

#### (2) 图像预处理技术。

在设计机器视觉系统时，我们优化选择光路系统、图像传感器和图像采集卡，以期获得高质量的场景图像信号。但是，图像的大小不一定正好满足用户的要求，而且，从图像传感、图像采集、输入到计算机这一整个过程中，因为多种干扰，图像质量或多或少有可能会发生退化 (如：灰度偏移、图像部分模糊等)。所以在图像后续分析和识别之前进行图像预处理就显得尤为必要。常用的图像预处理方法有：图像几何变换、灰度直方图变换、图像平滑和图像锐化等。

#### (3) 图像分割。

图像分割就是把图像分割成许多“有意义”的区域，使其后的图像分析、识别、理解等

更高层次的处理阶段所要处理的数据量大大减少，同时又保留有关图像有效的特征信息。

(4) 目标特征提取。

图像分割完成后，需要对图像中的目标抽取足以能够描述目标又能区分其他目标的特征，方便后续的特征分类、匹配以及图像理解。

(5) 特征分类与匹配。

根据从图像抽取的特征信息，把图像分成预定的类别。

(6) 图像理解。

给定一幅图像，图像理解程序不仅描述图像本身，而且描述和解释图像所代表的景物，以便对图像代表的内容作出决定。

(7) 行为决策。

根据图像理解，做出相应行为决策。

## 2. 机器视觉研究方法

计算机视觉技术集数字图像处理、数字信号处理、光学、物理学、几何学、应用数学、模式识别及人工智能等知识于一体，其应用已经涉及计算几何、计算机图形学、图像处理、机器人学等领域。因此，不可能将机器视觉的研究归结于用哪一种方法。但是，我们可以根据机器视觉的研究内容，将机器视觉系统划分成成像与捕捉、图像数据、图像特征和特征抽象不同层次的结构，如图 1-4 所示，在不同的层次上，选择相应的研究方法。

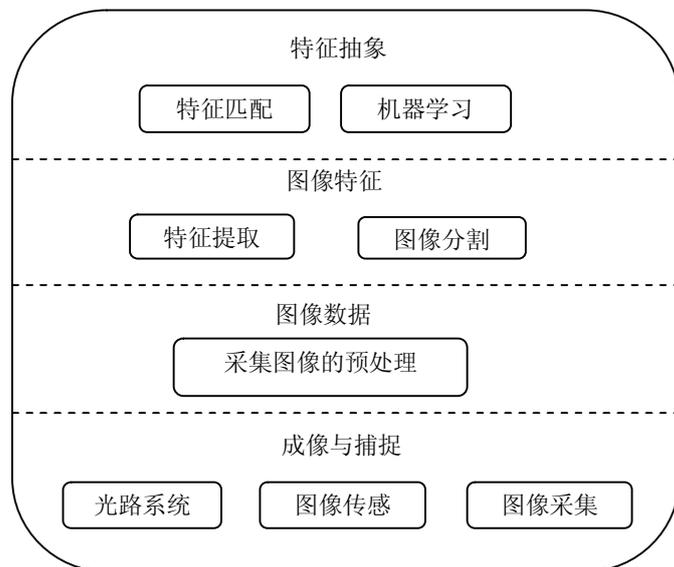


图 1-4 机器视觉的层次结构

(1) 成像与捕捉。

在成像与捕捉层次上，图像传感将受光路系统照射的场景转换成图像信号，经图像采集形成数字图像信号。由此可以看出，对这一层次的研究方法涉及物理学、光学、光电转换技术和电子技术。

## (2) 图像数据层。

在图像数据层次上，所处理的对象主要是像素级的数字信号，比如：图像的缩放、增强、降噪等。常用的研究方法包括几何的方法、数字信号处理的方法、直方图的方法、滤波的方法等。

## (3) 图像特征层。

在图像特征层次上，有两个主要任务，其一是图像分割；其二是特征提取。一般图像分割的方法主要有：边缘检测（Edge Detection）、边界跟踪（Edge Tracking）、区域生长、区域分离和聚合以及基于形态学的图像分割。

图像特征涵盖范围十分广泛，不仅包括形状、颜色、空间位置等与像素信息直接相关的底层图像特征，也包括更接近图像语义描述的各种统计学特征及频域纹理特征等。同时也有用于描述图像分布的全局特征及某目标区域的局部特征<sup>[9]</sup>。由于图像特征描述的多样性，导致了研究图像特征方法的多样性。一般图像特征描述的方法主要有：几何的方法（如：角点、拐点、特定的形状位置等特征点）、统计的方法（如：各阶矩）、频域的方法（如：傅里叶变换等）、多尺度的方法（如：小波变换）、人工神经网络的方法、人工智能的方法等。

## (4) 特征抽象层。

图像特征抽象层重点关注如何将所得到的图像特征解释为描述其内容的语义信息。常用的方法有模式识别和机器学习<sup>[9]</sup>。模板匹配、神经网络、支持向量机、AdaBoost 等都是和模式识别相关的技术；而姿势顺序词袋、视频特征词袋、基于稀疏矩阵的词袋、稀疏描述、AutoCode、卷积神经网络、受限玻尔兹曼机等都是和机器学习相关的技术。

### 3. 计算机视觉、机器视觉、图像识别之间的关系

计算机视觉、机器视觉、图像处理、图像识别与理解是彼此紧密关联的学科。从技术和应用领域的角度来看，机器视觉和计算机视觉有着相同的应用领域：

- (1) 过程控制，如：工业机器人以及流水线监控。
- (2) 导航，如：自主汽车或移动机器人。
- (3) 监测和跟踪，如：港口、码头、机场、银行等重点场所的视频监控和人数统计。
- (4) 机器视觉在医学中的应用。

在医学领域，机器视觉用于辅助医生手术（如：手术部位精确定位、医生实习培训等）和医学影像分析（如：主要利用数字图像处理技术、信息融合技术对 X 射线透视图、核磁共振图像、CT 图像进行分析，辅助医生诊断）。

## (5) 机器视觉在卫星遥感中的应用。

机器视觉技术被用于分析各种遥感图像，进行环境监测、地理测量，根据地形、地貌的图像和图形特征，对地面目标进行自动识别、理解和分类等。

(6) 自动检测，如：食品安全检测、印刷检测、制造业中表面检测、机器人定位以及“手-眼”系统、产品包装印刷质量的检测、半导体集成块封装质量检测等。

计算机视觉和机器视觉二者在实际应用中的侧重点不同。计算机视觉比较通用，而机器视觉侧重于检测和机器人视觉两个方面。

计算机视觉和机器视觉有着共同的基础理论和技术，即图像处理、图像识别与理解，二者并没有很清晰的界限，都是要从图像或图像序列中获取对场景的识别和理解。

计算机视觉是计算机科学的一个分支，而机器视觉是计算机视觉在系统工程中的一个工业应用。但是在获取高速处理速度上经常会使用特殊的图像处理硬件，这个速度是普通计算机所不能达到的。

由于机器视觉主要用于工业领域，所以对构成机器视觉系统的摄像机的拍摄速度、分辨率、机器视觉系统的处理速度以及鲁棒性和稳定性都提出了比较高的要求，不同于一般的计算机视觉系统。

## 1.3 统计机器学习与机器视觉

### 1.3.1 两者的关系

在介绍统计机器学习与机器视觉的关系之前，有必要对模式识别做一说明。模式识别（Pattern Recognition）是指对表征事物或现象的各种形式的（数值的、文字的和逻辑关系的）信息进行处理和分析，以对事物或现象进行描述、辨认、分类和解释的过程，是信息科学和人工智能的重要组成部分<sup>[10][11]</sup>。

计算机视觉（或机器视觉）是从图像处理和模式识别发展起来的，它的基础包括射影几何学、刚体运动力学、概率论与随机过程、图像处理、人工智能等理论。

机器学习正如前面介绍的那样，是研究计算机怎样模拟或实现人类的学习行为，以获取新的知识或技能，重新组织已有的知识结构使之不断改善自身的性能。它是人工智能的核心，是使计算机具有智能的根本途径，其应用遍及人工智能的各个领域，如专家系统、自动推理、自然语言理解、模式识别、计算机视觉、智能机器人等领域。

统计机器学习是机器学习发展到一定阶段的产物，是关于计算机基于数据构建的概率统计模型并运用模型对数据进行预测和分析的一门科学。

综观上述计算机视觉、模式识别、图像处理、机器学习和统计机器学习的描述，我们可以得出如下结论：

（1）要实现计算机视觉离不开图像处理，模式识别是图像特征匹配的有效方法之一，而模式识别是人工智能领域的一个重要分支，人工智能与机器学习密不可分，计算机视觉又应用服务于机器学习。这些环节，相辅相成，互相促进<sup>[10]</sup>。

（2）机器学习（统计机器学习）属于方法论的范畴，计算机视觉（机器视觉）是应用场景。解决机器视觉领域的问题会大量用到 ML 中的很多方法，但机器学习中的方法不仅局限于计算机视觉，还可以应用到数据挖掘、金融分析、人工智能等领域。举个具体例子，计算机视觉中的运动目标视觉跟踪。解决这个 CV 问题的方法中，有一大类方法属于判别型的方法，即把跟踪问题看成是在每一帧图像上目标与背景之间的二分类问题。跟踪过程中通过不断搜集更新正负样本，训练分类器，从当前待选目标中选取分类器响应最大的为该帧的跟踪结果。这里的二分类问题，就是典型的 ML 中的方法，例如，支撑向量机、朴素贝叶斯等<sup>[10]</sup>。

（3）本书主要研究统计机器学习在机器视觉中的应用。

### 1.3.2 两者的数学基础

机器视觉是一个涉及面非常广的学科，包括概率统计、各类优化方法、图论；一些研究方向（比如涉及物体运动的）还会涉及拓扑学、群论、矩阵优化；一些图像分割算法，会涉及微分方程等<sup>[12]</sup>。所以，学习和应用统计机器学习与机器视觉，应具备如下主要数学知识。

#### 1. 线性代数和统计学<sup>[13]</sup>

线性代数和统计学代表了机器学习中最主流的两类方法的基础。一种是以研究函数和变换为重点的代数方法，比如 Dimension Reduction、Feature Extraction、Kernel 等，一种是以研究统计模型和样本分布为重点的统计方法，比如 Graphical Model、Information Theoretical Models 等。它们侧重虽有不同，但常常是共同使用的，对于代数方法，往往需要统计上的解释，对于统计模型，其具体计算则需要代数的帮助。以代数和统计为出发点，继续往深处走，我们会发现需要更多的数学。

#### 2. 图论

由于它在表述各种关系方面的强大能力以及优雅的理论、高效的算法，越来越受到机器学习领域的欢迎。

#### 3. 微积分

微积分只是数学分析体系的基础。机器学习研究的大部分问题是在连续的度量空间进行的，无论代数还是统计，在研究优化问题的时候，对一个映射的微分或者梯度的分析总是不可避免的<sup>[14]</sup>。

#### 4. 傅里叶变换与小波变换

#### 5. 矩阵分析

## 参考文献

- [1] 李航. 统计学习方法. 北京: 清华大学出版社, 2012.
- [2] 王珏, 周志华, 周傲英. 机器学习及其应用. 北京: 清华大学出版社, 2006.
- [3] 王雪松, 程玉虎. 机器学习理论、方法及应用. 北京: 科学出版社, 2009.
- [4] 张燕平, 张铃. 机器学习理论与算法. 北京: 科学出版社, 2012.
- [5] <http://wenku.baidu.com>.
- [6] 刘建伟, 刘媛, 罗雄麟. 半监督学习方法. 计算机学报, 2015, 38 (8): 1592-1617.
- [7] 曹莹, 苗启广, 刘家辰, 高琳. AdaBoost 算法研究进展与展望. 自动化学报, 2013, 39 (6): 745-758.
- [8] 刘成君, 戴汝为. 计算机视觉研究的进展. 模式识别与人工智能, 1995, 38 (增): 59-71.
- [9] 徐志杰, 王晶等. 计算机视觉核心技术现状与展望. 西安邮电学院学报, 17 (6): 1-8.
- [10] <http://www.blogbus.com/shijuanfeng-logs/216968430.html>.
- [11] <http://blog.sina.com.cn/shijuanfeng>.
- [12] <http://www.zhihu.com/question/28949635/answer/42709647>.
- [13] <http://www.shookr.com/tech/426-2015-01-19-10-16-10>.