**第七章 机器学习**

**导读**

机器学习（Machine Learning）属于人工智能领域，其重点是开发用于自动获取知识的原理和技术。通过直接从现有数据库中提取知识，某些机器学习方法可以大大降低开发基于知识的软件成本。其他机器学习方法使软件系统可以在不需人工干预的情况下随时间改善其性能。人工智能是智能机器如计算机所执行的与人类智能有关的功能，如识别、判断、证明、学习等思维活动。这反映了人工智能学科的基本思想和内容，即人工智能是研究人类智能活动的规律的一门学科。

**学习目标**

1. 了解机器学习的概念以及研究目标
2. 熟知机器学习常用的方法
3. 了解机器学习的发展历史
4. 理解监督学习、无监督学习、半监督学习等常见机器学习方法

7.1 简述

现如今人工智能已经不再是一个小众化的研究课题了，全世界几乎所有的理工科类大学都在研究这门学科，甚至为此设立专门的研究机构。越来越多的学习计算机、自动控制和软件工程专业的本科生或研究生将人工智能作为自己的研究方向。在科学家的不懈努力下，如今计算机与原来相比已经变得十分聪明了，某些时候计算机已经可以完成原来只属于人类的工作，且其高速性和准确性是人类远不可及的。机器学习是一门涉及多领域的交叉学科，其包含高等数学、统计学、概率论、凸分析、逼近论等多门学科。该学科专门研究计算机应如何模拟并实现人类的学习行为，以获取人类所不了解的新的知识，并使计算机能够使用已有的知识或经验不断改善自身的性能以得到更加精确的知识。



图1 机器学习 [1]

当前，属于机器学习范畴的基于人工神经网络的深度学习技术是人工智能方向最热门的研究领域，被Google，Facebook，IBM，百度，NEC以及其他互联网公司广泛使用，用来进行图像和语音识别。人工神经网络从20世纪80年代兴起，经过科学家们的不懈努力，相关算法不断被优化并处于持续改进和创新之中，同时也受益于计算机技术的快速发展，现在科学家可以通过GPU模拟建立超大型的人工神经网络；互联网行业的快速发展，为深度学习提供了百万级的样本进行训练，在上述三个因素共同作用下现在的语音识别技术和图像识别技术能够达到90%以上的准确率。



图2 语言识别[2]

机器学习的研究目标有三个方向，一个方向是以模拟人类的学习过程出发，试图建立学习的认识生理学模型，这个方向与认知科学的发展密切相关。第二个方向是基础研究，发展各种适合机器特点的学习理论，探讨所有可能的学习方法，比较人类学习与机器学习的异同与联系。第三个方向是应用研究，建立各种实用的学习系统或知识获取辅助工具，在人工智能科学的应用领域，机器人系统，专家系统等建立自动获取知识系统，积累经验，完善知识库与控制知识，进而能使机器的智能水平像人类一样。

机器学习的方法，可采用模拟人类学习的方法，也可根据机器自身的特点采用新的方法。更重要的是将两种方法结合起来，人类的知识和才能并非天赋的，生来就有的，而是后天不断学习的结果，人的学习过程就是一个认识过程，这个过程离不开人类的社会环境，实践、知识、认识三者相互反复作用，构成了认识论的总体模型，也为建立学习模型提供了依据。同时，人的学习具有生物学特性，儿童时期的学习最为基础，然而，人类的学习至少存在两大缺陷，学习过程非常缓慢和无法复制。机器的学习应充分运用人类学习方法上的研究成果，也应根据机器自身的特点，如快速，存储量大，易复制等优点，研究出适合机器特点的学习方法。

近年来，人们开始关注使用bagging、boosting或bootstrapping等技术来生成和组合不同但同质的分类器。它们基于重复生成相同类型的模型，而不是不断进化的训练数据。这些方法能够减少模型方差。由于该方法对数据的扰动具有稳定性，因此不能应用于基于实例的学习周期。

7.2 机器学习常用的方法

决策树学习是一种利用决策树逼近离散函数的方法。树的节点中是属性，叶子中是离散函数的值。决策树可以用一组if-then规则重写。树学习方法是一种流行的归纳推理算法，主要用于各种分类任务(如诊断病例)。在树的生成中，通常使用熵作为属性的信息增益度量。最著名的方法是ID3、C4.5等。

神经网络学习方法提供了一个鲁棒的方法来逼近重值、离散值和向量值函数。众所周知的反向传播算法使用梯度下降来调整网络参数，使其最适合输入-输出对的训练集。这种方法受到神经生物学的启发，它模仿大脑的功能，在那里许多神经元相互连接，实例由许多输入-输出对表示。神经网络学习对训练数据中的错误具有良好的鲁棒性，并已成功地应用于语音识别、人脸识别等问题。

贝叶斯推理提供了一种概率推理方法。贝叶斯推理为直接操纵概率的学习算法提供了基础，也为分析其他算法的操作提供了框架。贝叶斯学习算法为假设计算明确的概率，如朴素贝叶斯，是某些类型中学习问题最实用的方法之一。在许多情况下，贝叶斯分类器与其他机器学习算法具有竞争力。例如，对于学习分类文本文档，朴素贝叶斯分类器是最有效的分类器之一。

强化学习解决了任务如何学会选择最佳的行动来达到它的目标。每当代理在它的环境中执行一个动作时，训练者可能会提供奖励或惩罚来表明结果状态的便利性。例如，当代理被训练去玩一个游戏，那么当游戏获胜时，训练者可能提供一个积极的奖励，当游戏失败时提供一个消极的奖励，而在其他状态时则提供零奖励。代理的任务就是从这种延迟的奖励中学习，选择一系列能产生最大累积奖励的行为。从延迟奖励中获得最优控制策略的算法称为Q-learning。该方法可以解决学习控制移动机器人、学习优化工厂操作、学习规划治疗程序等问题。

归纳逻辑编程源于从例子中学习概念，是一种相对简单的归纳形式。概念学习的目的是从一组预先分类的例子中发现一组具有高预测能力的分类规则。ILP理论是基于一阶谓词演算的证明理论和模型理论。归纳假设的形成，其特点是采用了逆分解、相对最小泛化、逆暗示和逆蕴涵等技术。该方法可用于从训练数据集创建逻辑程序，最终的程序能够生成该数据。创建逻辑程序非常依赖于任务的复杂性。在许多情况下，如果没有对最终程序的许多限制，这种方法是不可用的。ILP成功地应用于数据挖掘，用于在大型数据库中寻找规则。

基于案例的推理(Case-Based Reasoning, CBR)是一种惰性学习算法，它通过分析相似的实例来分类新的查询实例，同时忽略与查询有很大差异的实例。此方法在实例内存中保存以前的所有实例。实例可以用值、符号、树、各种层次结构或其他结构来表示。这是非泛化的方法。CBR的工作周期是:案例检索-重用-解决方案测试-学习。这种方法受到生物学的启发，具体地说，是人类利用旧的类似情况的知识进行推理的结果。这种学习方法也被称为类推学习。CBR范式涵盖了一系列不同的方法。广泛使用的是基于实例的推理(Instance-Based Reasoning, IBR)算法，它与一般的CBR主要区别在于表示实例。实例的表示很简单，通常是数值或符号值的向量。基于实例的学习包括k近邻(k-NN)和局部加权回归(LWR)方法。

支持向量机(SVM)是近年来非常流行的分类和优化方法。SVM是由Vapnik等人在1992年引入的。这种方法结合了两个主要思想。第一个是最优线性边缘分类器的概念，它构造了一个分离超平面，使到训练点的距离最大化。第二个是内核的概念。核函数的最简单形式是计算两个训练向量的点积的函数。内核在特征空间中计算这些点积，通常没有明确地计算特征向量，而是直接对输入向量进行操作。使用特征变换将输入向量重新表示为新的特征时，即使新的特征空间具有更高的维数，也要在特征空间中计算点积。所以线性分类器不受影响。Margin maximization提供了一个有用的权衡分类精度，这很容易导致训练数据的过拟合。支持向量机很好地适用于属性数量相对于训练示例数量较大的学习任务。

遗传算法提供了一种学习方法，其动机类似于生物进化，寻找一个合适的假设。当前种群的成员通过选择、交叉和突变等操作产生下一代种群。在每一步中，一个称为当前种群的假设集合被更新，方法是用当前最适合的假设的后代替换一部分种群。遗传算法已成功地应用于各种学习任务和优化问题。例如，遗传算法可用于其他ML方法，如神经网络或基于实例的推理，以获得最佳参数设置。



图3 遗传算法[3]

7.3 机器学习的发展

从1950年代发展到现在，经历了推理期、知识期、学习期。

（1）推理期。在推理期，只要给予机器逻辑推理能力，机器学习就具备智能。同时期产生了A.Newell和H.Simon的Logic Theorist程序和General Problem Solving 程序。但是这时的逻辑推理机器，不能满足人工智能研究的需求。

（2）知识期。在1970年代人工智能的发展进入了知识期，知识期即由人把知识总结出来再教给计算机。这一时期发展了大量专家系统，人工智能在众多应用领域取得了大量成果。随着知识系统的发展，专家系统越来越复杂，人们发现这种人工总结知识教给计算机做法 非常困难的，专家系统迎来了知识工程瓶颈。于是，人工智能迎来了学习期，即机器自己学习知识，也就是机器学习。

（3）学习期。追溯机器学习的第一次出现，是在1950年图灵关于图灵测试的文章中。其后在1950年代中后期基于神经网络的连接主义（Connectionism）学习开始出现， 产生了F.Rosenbatt的感知机（Perceptron）， B.Widrow 的Adaline等。此时，基于decision theory（决策理论）为基础的学习技术和强化学习技术发展迅速。随后在1980年代，机器学习成为一个独立学科，各种机器学习技术百花齐绽。1980年第一届机器学习研讨会IWML在美国卡耐基梅陇大学 （Carnegie Mellon University）举行。同期各种机器学习专业期刊，有关机器学习的专辑文章雨 后春笋般的出现。1980年代，R.S.Michalski等人和E.A.Feigenbaum等人在著名的《人工智能手册》 中把机器学习划分为如下部分：



图4 图灵测试[4]

①R.S.Michalsk 等人提出从样例中学习，在问题求解和规划中 学习，通过观察和发现学习，从指令中学习。

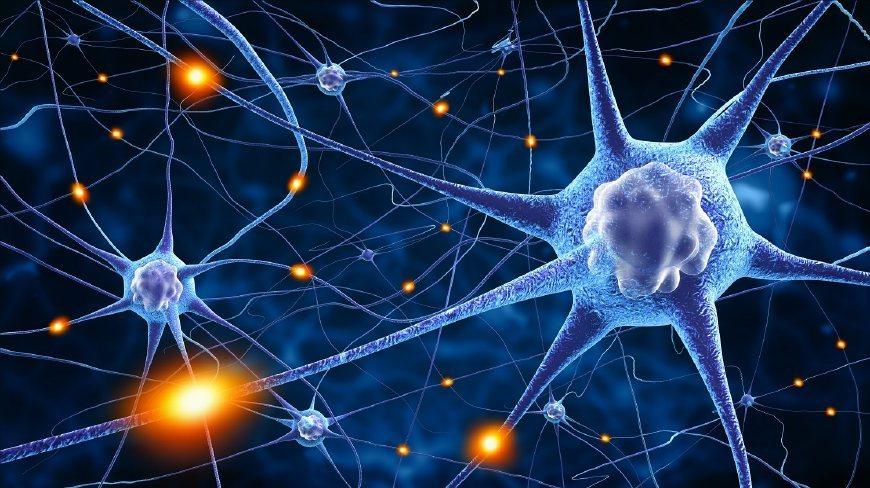
②E.A.Feigenbaum等人提出归纳学习，机械学习， 类比学习，示教学习。从样例中学习对应广义的 “归纳学习”，即从训练样本中归纳出学习结果。

（1）决策树和基于逻辑的学习。1980年代，以决策树理论和基于逻辑的学习得到快速发展。

① 决策树理论，信息熵作为变量出现的期望值，信息熵越小系统越有序，信息熵的最小化成为决策树的目标。

②基于逻辑的学习，其中归纳逻辑程序设计（ILP）是使用一阶逻辑来进行知识表示，通过修改和扩充逻辑表达式来完成对数据的归纳。ILP引入了逻辑表达式嵌套和函数，于是机器学习具备了强大的表达能力。但是同时带来了巨大的挑战， 如给定P为一元谓词，f为一元函数，可以组成 P(x)、P(f(x))等无限个，所以候选原子公式在规则学习过程中有无限个。在增加规则时，造成的 复杂度太高，而无法进行有效的学习。1990年代中后期这方面的研究相对进入低潮。

（2）基于神经网络的连接主义学习。1990年代前期，基于神经网络的连接主义学习发展成为主流技术。1986 年，D.E.Rumelhart等人重新发明了著名的BP算法，产生了深远的影响，由于BP算法十分高效，使它在很多现实问题上发挥作用。在连接主义学习过程中，存在观点的多样性，可能会包含大量的参数，参数的选择全靠经验判断，没有理论依据，人工调整参数可能导致误差，学习结果将会差距很大。

  
图5 神经网络[5]

（3）统计学习支持向量SVM。1990 年代中期，统计学习迅速崛起，支持向量机SVM成为其代表性技术。到2000年代，深度学习快速发展，深度学习狭义的解释就是多层神经网络。计算机硬件处理技术与数据存储技术的发展，给深度学习带来了更加广阔的发展和应用，深度学习技术快速席卷了整个人工智能领域。处理器技术的发展也使得人工智能能够在数据和高运算能力下发挥它的作用。

7.4 监督学习

7.4.1监督学习概述

监督学习也称有导师的学习，指在训练期间有一个外部导师告诉网络每个输入向量的正确的输出向量。学习的目的就是减少网络产生的实际输出向量和预期输出向量之间的差异。这一目标是通过逐步调整网络内的权值实现的，反向传播算法能够决定权值要改变多少。对于这种学习，网络在能执行工作前必须训练。当网络对于给定的输入能产生所需要的输出时，就认为网络的学习和训练已经完成。由此可以看到，监督学习的成分主要有：实际输出向量；预期输出向量；实际输出向量和预期输出向量之间存在的差异等。这样，就可以具体分析某一学习活动，根据其所包含的成分，从而推断其是否是监督学习。

从给定的训练数据集中学习出一个函数（模型参数），当新的数据到来时，可以根据这个函数预测结果。监督学习的训练集要求包括输入输出，也可以说是特征和目标。训练集中的目标是由人标注的。监督学习就是最常见的分类（注意和聚类区分）问题，通过已有的训练样本（即已知数据及其对应的输出）去训练得到一个最优模型（这个模型属于某个函数的集合，最优表示某个评价准则下是最佳的），再利用这个模型将所有的输入映射为相应的输出，对输出进行简单的判断从而实现分类的目的。也就具有了对未知数据分类的能力。监督学习的目标往往是让计算机去学习我们已经创建好的分类系统（模型）。

监督学习是训练神经网络和决策树的常见技术。这两种技术高度依赖事先确定的分类系统给出的信息，对于神经网络，分类系统利用信息判断网络的错误，然后不断调整网络参数。对于决策树，分类系统用它来判断哪些属性提供了最多的信息。常见的有监督学习算法：回归分析和统计分类。最典型的算法是KNN和DT。

7.4.2 常见监督学习算法

一. K-近邻算法（k-Nearest Neighbors，KNN）

K-近邻是一种分类算法，其思路是：如果一个样本在特征空间中的k个最相似(即特征空间中最邻近)的样本中的大多数属于某一个类别，则该样本也属于这个类别。K通常是不大于20的整数。KNN算法中，所选择的邻居都是已经正确分类的对象。该方法在定类决策上只依据最邻近的一个或者几个样本的类别来决定待分样本所属的类别。

算法的步骤为：

（1）计算测试数据与各个训练数据之间的距离；

（2）按照距离的递增关系进行排序；

（3）选取距离最小的K个点；

（4）确定前K个点所在类别的出现频率；

（5）返回前K个点中出现频率最高的类别作为测试数据的预测分类。

二. 决策树（Decision Trees ，DT）

决策树是一种常见的分类方法，其思想和“人类逐步分析比较然后做出结论”的过程十分相似。决策过程和下图1类似。

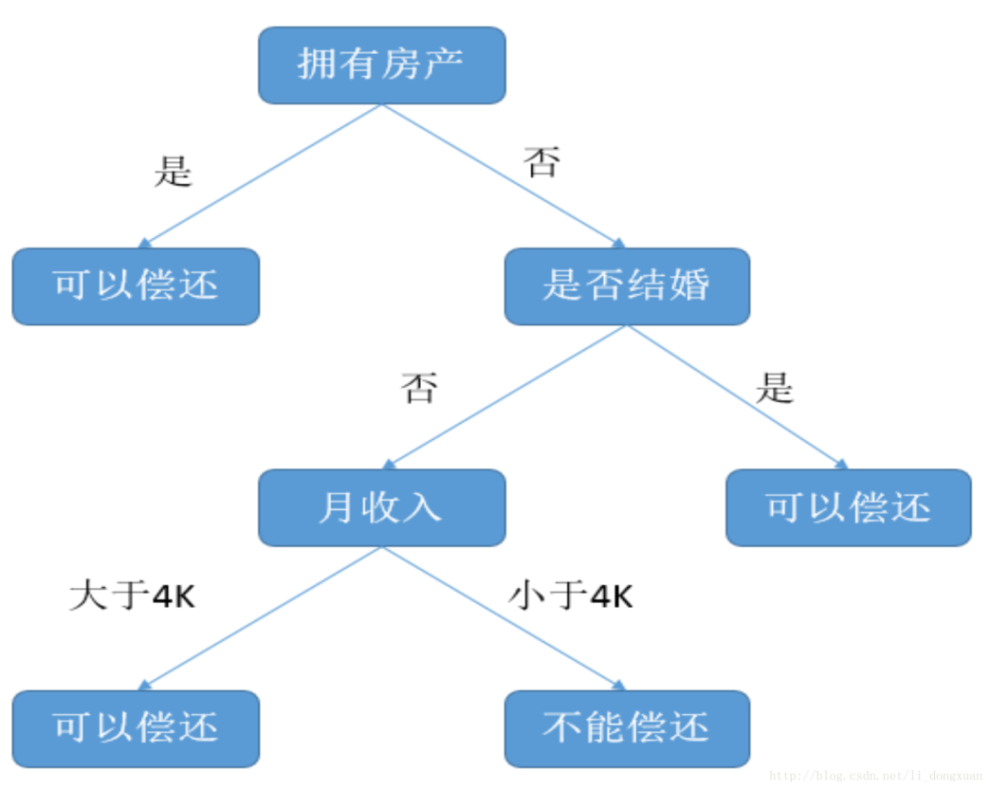


图6 决策过程

决策树是一个树结构（可以是二叉树或非二叉树）。其每个非叶节点表示一个特征属性上的测试，每个分支代表这个特征属性在某个值域上的输出，而每个叶节点存放一个类别。使用决策树进行决策的过程就是从根节点开始，测试待分类项中相应的特征属性，并按照其值选择输出分支，直到到达叶子节点，将叶子节点存放的类别作为决策结果。

不同于贝叶斯算法，决策树的构造过程不依赖领域知识，它使用属性选择度量来选择将元组最好地划分成不同的类的属性。所谓决策树的构造就是进行属性选择度量确定各个特征属性之间的拓扑结构。

7.5 无监督学习

7.5.1无监督学习概述

无监督学习是指无教师监督的学习过程， 即其神经网络学习样例是不带类别标号。它在广义上可分成两个子类，一个是强化学习，一个是自组织学习。计算机的出现解决了人脑的逻辑思维问题，而且功能远超人类， 但在当前计算机很难代替人脑独特的形象思维和智能。计算机与人在物体识别方面，花费时间和正确率有很大的差距。这种差距引起了科学家的关注，他们研究用工程方法来实现或近似模拟人类智能，从而出现了人工智能及神经网络，但这些都有一定的缺陷。然后，一种结合人对智能和对世界的理解的，模拟人眼视觉和人类智能的新的智能处理方法—无监督学习出现了。最常见的应用场景是用于故障诊断，如图7。

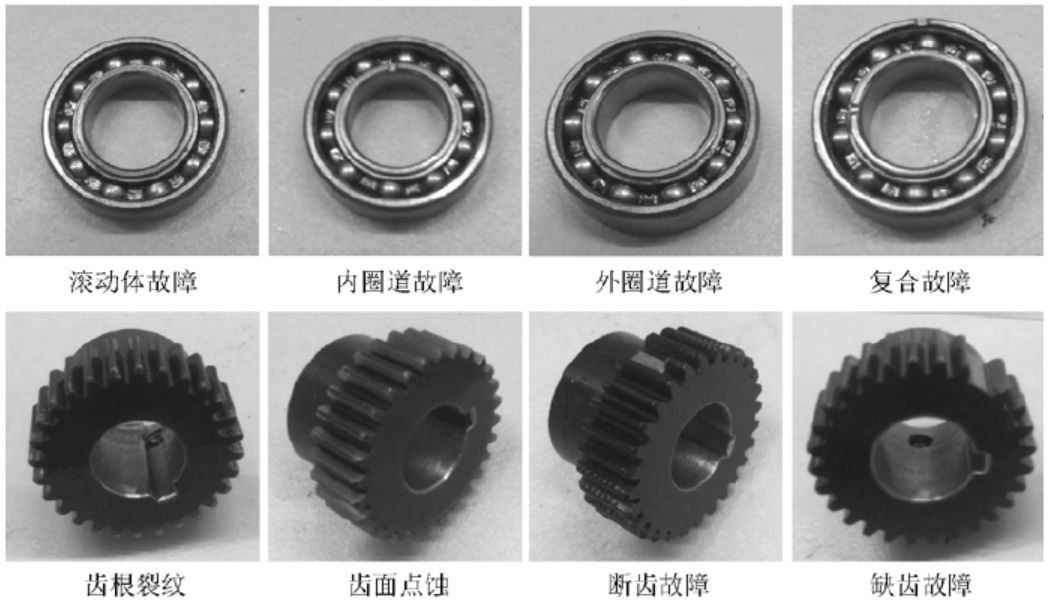


图7 机械故障分类

人工神经网络是由许多人工神经单元组成的，每个单元能依照“映射”并行计算，同时可通过样例学习。 样例的输出是已知的，又称有监督学习。反之，有一种 方法是根据神经元自己所处的状态学习的。假设两个神经元输出分别为xi、yi，wij（=xiyi）为它们之间的连接权， 如果两个神经元都处于兴奋的状态，则wij 值高；若都处于抑制状态，即wij＝0；这种方法称为无监督的学习。

无监督学习也是自组织学习系统，学习没有外部监督。自组织学习的训练目标不是产生一个分类系统， 而是对那些正确或错误的行为做出激励或处罚。所以 必须网络表达质量的任务度量，让学习根据这个度量 来最优化网络。为了完成自组织学习，我们可以使用竞争性学习规则。 所谓强化学习是指在此过程中，并不直接告诉机器要做什么或采取哪些行动，而是机器通过一些已有的不确定的信息来进行学习，做出最优的策略，得到最多的奖励来自己发现。机器所响应的动作的影响不仅是即刻得到的奖励，还影响接下来的动作和一连串的奖励。 强化学习的目标是将代价函数最小化。

7.5.2 自组织学习的相关形式

自组织映射（SOM）是基于竞争学习的，在SOM里， 神经元被放置在网格节点上，这个网络通常是一维或者是二维的，更高维的映射不常见。在竞争学习过程中， 用不同输入模式刺激，网络选择性地调整，形成对不同输入特征的机系。

自组织目标函数的互信息：在输入和输出随机过程之间的香农互信息具有一些唯一的性质，这些性质使其可作为自组织学习的目标函数，从而被优化。

有以下 4 种自组织原则：

（1）Infomax 原则，其包含了最大化神经网络的多维输入和输出向量之间的互信息，这一原则制定了自组织模型和特征映射的开发框架。

（2）最小冗余原则，这基本上是另一种最大化网络的输入和输出之间的互信息导致冗余最小化的方法。

（3）Imax 原则，这是最大化一对神经网络的单一输出之间的互信息，这对神经网络是由两个空间位移多维输入向量所驱动的。该原则非常适合于图像处理，目标是发现带噪声传感的输入在空间和事件上表现的相干性。

（4）Imin 原则，这是最小化一对神经网络的单一输出之间互信息，这对神经网络是由两个空间位移多维输入向量所驱动的。该原则在图像处理中的应用目标在于最小化同一环境中两幅相关图像之间的空间时间相干，图像是由具有正交性质的一对传感器获得的。

自组织学习的另一个类别是统计力学。统计力学作为优化技术表示和机器学习的数学基础。有以下3 种 模拟算法：

（1）Metropolis 算法，这是MCMC（Markov chain Monte Carlo）针对未知概率分布上的模拟。

（2）模拟退火，这是一个动态的过程，利用“高温时观察到系统的总特点，低温时出现细节特征” 来避免局部极小值的一种优化算法。

（3）Gibbs 抽样，它产生一个带Gibbs分布作为平衡分布的马尔科夫链。与 Metropolis算法不同，与Gibbs抽样器相关的转移概率不是静态的。

7.5.3 无监督学习的应用

乳腺癌早期诊断是很困难的，一般摄影只能观察几个病变像素，易被作为杂躁而忽视。利用两个不同的波段红外感应相机同时拍摄两幅图像，肿瘤在不同的 生长阶段，血管血液成分有不同的比例，从而呈现不同的红外特征。

不同波长的红外图像从两个通道输入神经网络，用 S1，S2 表示两幅红外图像中单像素的值，A 和 B 表示混合传递函数的矢量，让二维向量X=S1A+S2B，如何寻找 两个W1 和W2，获得S1 和S2。一个方法是让W1 与A正交、 W2 与 B 正交，即 S′2=W1·X=S2W1·B，S′1=W2·X=S1W2·B， 这样得到S′2 只与S2 有关，而 S′1 只与S1 有关。这样 对两幅乳房的图像进行逐个像素的处理，很快就可得到 确诊。这种采用正交向量对消元的无监督学习的方法， 称独立元分析法。



图7 监督学习与医疗诊断[6]

7.6 半监督学习

7.6.1　半监督学习概述

半监督学习（Semi-Supervised Learning, SSL）是机器学习（Machine Learning，ML）领域中的研究热点，已经被应用于解决实际问题，尤其是自然语言处理问题。SSL被研究了几十年，国内外涌现出大量关于该领域的研究工作，研究人员在这个问题上已 经取得了显著的进步。

在许多机器学习的实际应用中，如网页分类、文本分类、基因序列比对、蛋白质功能预测、语音识别、自然语言处理、计算机视觉和基因生物学，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，如图2所示，半监督学习避免了数据和资源的浪费，同时解决了监督学习的模型泛化能力不强等问题。

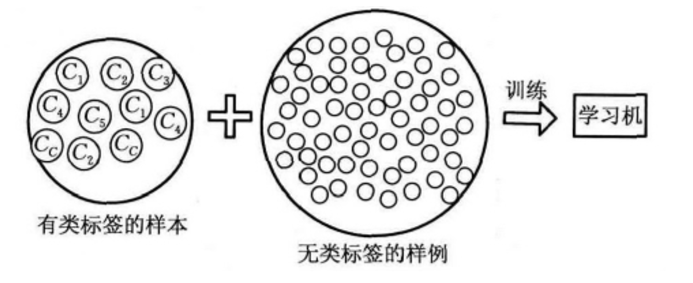


图8 半监督学习示意

7.6.2 半监督学习依赖的假设

半监督学习的成立依赖于模型假设，当模型假设正确时，无类标签的样例能够帮助改进学习性能。半监督学习依赖的假设有以下3个：

（1）平滑假设（Smoothness Assumption）。位于稠密数据区域的两个距离很近的样例的类标签相似，也就是说，当两个样例被稠密数据区域中的边连接时，它们在很大的概率下有相同的类标签；相反地，当两个样例被稀疏数据区域分开时，它们的类标签趋于不同。

（2）聚类假设（Cluster Assumption）。当两个样例位于同一聚类簇时，它们在很大的概率下有相同的类标签。这个假设的等价定义为低密度分离假设（Low Sensity Separation Assumption），即分类决策边界应该穿过稀疏数据区域，而避免将稠密数据区域的样例分到决策边界两侧。

（3）流形假设（Manifold Assumption）。将高维数据嵌入到低维流形中，当两个样例位于低维流形中的一个小局部邻域内时，它们具有相似的类标签。 许多实验研究表明当半监督学习不满足这些假设或模型假设不正确时，无类标签的样例不仅不能对学习性能起到改进作用，反而会恶化学习性能，导致半监督学习的性能下降．但是还有一些实验表明，在一些特殊的情况下即使模型假设正确，无类标签的样例 也有可能损害学习性能．例如，Shahshahani和Landgrebe通过实验证明了如何利用无类标签的样例帮助减轻休斯现象（Hughes Phenomenon）（休斯现象指在样例数量一定的前提条件下，分类精度随着特征维数的增加先增后降的现象），但是同时实验中也出现了无类标签的样例降低学习性能的情况。Baluja用朴素贝叶斯分类器和树扩展朴素贝叶斯（Tree Augmented Na ve Bayesian，TAN）分类器得到很好的分类结果，但是其中也存在无类标签的样例降低学习性能的情况。 Balan和Blum提出容许函数使分类器能够很好的服从无类标签的样例的分布，但是这种方法同样会损害学习性能。

　7.6.3 半监督学习的分类

　半监督学习按照统计学习理论的角度包括直推（Transductctive）半监督学习和归纳（Inductive）半监督学习两类模式。直推半监督学习只处理样本空间内给定的训练数据，利用训练数据中有类标签的样本和无类标签的样例进行训练，预测训练数据中无类标签的样例的类标签；归纳半监督学习处理整个样本空间中所有给定和未知的样例，同时利用训练数据中有类标签的样本和无类标签的样例，以及未知的测试样例一起进行训练，不仅预测训练数据中无类标签的样例的类标 签，更主要的是预测未知的测试样例的类标签。

从不同的学习场景看，半监督学习可分为4大类：

（1）半监督分类 （Semi-Supervised Classification）。在无类标签的样例的帮助下训练有类标签的样本，获得比只用有类标签的样本训练得到的分类器性能更优的分类器，弥补有类标签的样本不足的缺陷，其中类标签yi 取有限离散值 yi∈{c1, c2,…,cj} ∈N 。

（2）半监督回归（Semi-Supervised Regresion）。在无输出的输入帮助下训练有输出的输入，获得比只用有输出的输入训练得到的回归器性能更好的回归器，其中输出 yi取连续值yi ∈R。

（3）半监督聚类（Semi-Supervised Clustering）。在有类标签的样本的信息帮助下获得比只用无类标签的样例得到的结果更好的簇，提高聚类方法的精度。

（4）半监督降维（Semi-Supervised Dimensionality Reduction）。在有类标签的样本的信息帮助下找到高维输入数据的低维结构，同时保持原始高维数据和成对约束（Pair-Wise Constraints）的结构不变，即在高维空间中满足正约束（Must-Link Constraints） 的样例在低维空间中相距很近，在高维空间中满足负约束（Cannot-Link Constraints）的样例在低维空间中距离很远。

为便于更加清晰地介绍各种半监督学习方法，这里按照图3对各种半监督学习方法进行归类。

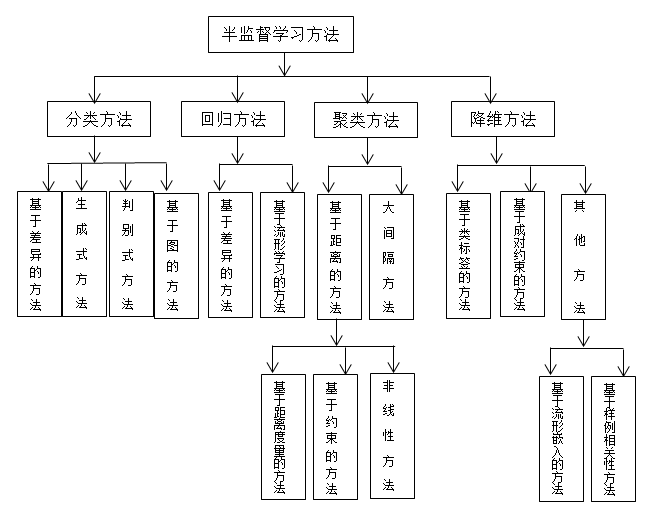


图8　半监督学习方法结构

**课后思考题**

1. 机器学习的主要研究内容是什么？
2. 机器学习常用的方法有哪些？
3. 机器学习的发展经历了哪些阶段？
4. 举例说明监督学习、无监督学习、半监督学习的区别。
5. 阐述监督学习、无监督学习、半监督学习各自的优缺点。

**参考文献**

[1] <https://developer.aliyun.com/article/179926>

[2]<http://www.pinggu.org/index.php?m=content&c=index&a=show&id=4240665&page=xy>

[3] <https://www.sohu.com/a/231085487_163476>

[4] <https://m.sohu.com/a/39315509_289427>

[5] <https://www.da-quan.net/so-yysdbcsqq.html>

[6]<http://www.elecfans.com/soft/yiliaodianzi/2019/0117/854237.html>