

1 这一节我们概述下主要算法的类别

2 数据处理系统：算法

大数据处理算法基本上包括数据挖掘算法和机器学习算法。

机器学习算法包括监督学习、无监督学习、半监督学习和强化学习。

监督学习是一种使机器能够根据输入机器的相关数据对对象、问题或情况进行分类的方法。

机器被重复地输入诸如物体、人或情况的特征、图案、尺寸、颜色和高度等数据，直到机器能够进行精确的分类。

监督学习是一种应用于现实场景的流行技术或概念。

监督学习用于提供产品推荐、根据客户数据细分客户、根据以前的症状诊断疾病以及执行许多其他任务

无监督学习是一种方法，用来使机器分类有形的有形的和无形的对象，而不提供机器任何关于对象的先验信息。

机器需要分类的东西多种多样，比如顾客的购买习惯、细菌的行为模式和黑客攻击。

无监督学习背后的主要思想是让机器接触大量不同的数据，并允许它从数据中学习和推断。

然而，首先必须对机器进行编程，使其能够从数据中学习。

在监督学习的范围内有回归(线性回归和多项式)，分类(逻辑回归，支持向量机，神经网络和决策树)和深度学习(CNN, RNN)。

在无监督学习的范围内，有聚类(如 K-means)、降维(PCA)和异常检测。

半监督学习包括自我训练和低密度分离模型。

强化学习包括动态规划和蒙特卡罗方法。

3 机器学习是对统计模型和算法的科学研究，计算机系统使用这些模型和算法在没有明确指令的情况下执行任务。基本上，机器学习允许计算机在不需要明确编程的情况下进行学习。

机器学习是一个综合性的领域，它涉及机器学习操作的各种功能，如聚类、分类和预测模型的开发。

在一般的编程场景中，您必须向计算机提供指令，以便它为您提供输出。

然而，在机器学习算法的帮助下，你可以训练你的计算机为你提供需要给出指令的输出。

机器学习算法能够在数据的帮助下做到这一点。使用输入到系统的数据，训练机器学习算法为用户提供输出。

强化学习(RL)是一种机器学习技术，它使智能体能够在交互式环境中通过尝试和错误学习，使用来自自身行动和经验的反馈。

尽管监督学习和强化学习都使用输入和输出之间的映射，但与提供给代理的反馈是执行任务的正确行动集的监督学习不同，强化学习使用奖励和惩罚作为积极和消极行为的信号。

与无监督学习相比，强化学习在目标上有所不同。

非监督学习的目标是找到数据点之间的相似点和不同点，而强化学习的目标是找到一个合适的行为模型使代理的总累积奖励最大化。

下图展示了强化学习模型所涉及的基本思想和元素。

4 深度学习

深度学习是一个新兴的领域，占据了机器学习的更广泛的领域。

深度学习最著名的是它的神经网络，如循环神经网络 RNN，卷积神经网络 CNN 和深度信念网络。

虽然其他机器学习算法使用“统计分析技术”进行模式识别，但深度学习是以人脑的神经元为模型的。

它们模仿了人类大脑的结构和功能。为了理解深度学习，我们必须了解人体神经系统是如何工作的。

众所周知，我们的神经系统是由神经元组成的。这些神经元能够捕捉传输到我们身体的信息。

这些神经元有能力随着时间的推移学习信息。这种“学习”的原理也被人工神经网络所利用。

任何深度神经网络都由三种类型的层组成:输入层:隐藏层和输出层

这两者有什么区别?

简单地说，机器学习和深度学习都模仿了人类大脑的学习方式。

因此，它的主要区别在于每种情况下使用的算法类型，尽管深度学习与人类学习更相似，因为它与神经元一起工作。机器学习通常使用决策树和深度学习神经网络，它们更进化。此外，两者都可以在监督或无监督的情况下学习。

5 这里对比下深度学习和机器学习

从表中可以看出，DL 需要大量的数据，ML 可以在较少的数据上进行训练，

DL 可以提供较高的准确率，而 ML 给出的准确率较低

;DL 训练时间更长，而 ML 训练时间更短，

L 要求 GPU 进行适当的训练，ML 在 CPU 上进行训练，

DL 可以通过各种不同的方式进行调优，但 ML 的调优能力有限。

6 数据挖掘算法

数据挖掘算法可分为 4 组:分类、聚类、相关分析和异常检测。

图中显示了排名前 12 位的数据挖掘算法，包括支持向量机、决策树、神经网络、ID3 算法、C4.5 算法、k 最近邻、Naïve 贝叶斯算法等。

7 这一节我们简要讨论了主要算法的几种主要类别，今天的学习就到这里，谢谢大家