

定义

(1) 机器学习是一门人工智能的科学,该领域的主要研究对象是人工智能,特别是如何在经验学习中改善具体算法的性能。

(2) 机器学习是对能通过经验自动改进的计算机算法的研究。

(3) 机器学习是用数据或以往的经验,以此优化计算机程序的性能标准。

(我的定义:人工智能是用【我们一直的技术】,来模拟【一个用【我们未知的(或不完全已知的)技术】创造的系统】(即所谓的合理行动agent))

(实际上人工智能的目标确实有很多流派,比如合理行动/合理思考/像人一样行动/像人一样思考)

分类

基于学习策略

1. 模拟人脑(符号学习、神经网络学习)

(这与“我的定义”描述的比较符合。但是我的定义其实不是这个意思。我的定义中,模拟的途径是任意的,只要模拟的结果)

2. 直接采用数学方法(统计机器学习)

(这个大概可以归为合理行动agent)

基于学习方法

1. 归纳学习
2. 演绎学习
3. 类比学习
4. 分析学习

基于学习方式

1. 监督学习

(感觉这个有点依靠人力,但是机器学习方法不也是人力造出来的吗?区别只在有没有自动化)

2. 无监督学习(发现学习、聚类学习、竞争学习)

3. 强化学习:以奖惩信号作为输入,以统计和动态规划技术为指导的一种学习方法。

(我比较支持这个。我觉得人类的理性就是在奖惩信号中发展起来的。至于感性,对于目前的机器学习,我们大部分情况下不需要)

基于数据形式

1. 结构化学习:以结构化数据为输入,如神经网络学习、决策树学习、规则学习。
2. 非结构化学习:亦非结构化数据为输入,如案例学习、类比学习、解释学习、各种挖掘

基于学习目标

1. 概念学习：学习的目标和结果为概念，或者说是为了获得概念的学习。典型的概念学习主要有示例学习。
2. 规则学习：学习的目标和结果为规则，或者为了获得规则的学习。典型规则学习主要有决策树学习。
3. 函数学习：学习的目标和结果为函数，或者说是为了获得函数的学习。典型函数学习主要有神经网络学习。
4. 类别学习：学习的目标和结果为对象类，或者说是为了获得类别的学习。典型类别学习主要有聚类分析。
5. 贝叶斯网络学习：学习的目标和结果是贝叶斯网络，或者说是为了获得贝叶斯网络的一种学习。其又可分为结构学习和多数学习。

常见算法以及各自优点总结

注：详见百度百科：机器学习

决策树算法

结构简单，处理数据效率较高

朴素贝叶斯算法

对于高维数据点或大量数据点表现较好

支持向量机算法

输入空间、提高维度从而将问题简短化，使问题归结为线性可分的经典解问题

应用于垃圾邮件识别、人脸识别等问题

随机森林算法

可以产生高精度的分类器，并能够处理大量的变数，也可以平衡分类资料集之间的误差。

人工神经网络算法

此方法有很强的泛化能力和非线性映射能力，可以对信息量少的系统进行模型处理。从功能模拟角度看具有并行性，且传递信息速度极快。

Boosting与Bagging算法

它可以将弱学习算法提高为强学习算法，可以应用到其它基础回归算法，如线性回归、神经网络等，来提高精度。

关联规则算法

用规则去描述两个变量或多个变量之间的关系，以客观反映数据本身性质。

EM（期望最大化）算法

它是对本身受限制或比较难直接处理的数据的极大似然估计算法。

深度学习

学习样本数据的内在规律和表示层次，这些学习过程中获得的信息对诸如文字，图像和声音等数据的解释有很大的帮助。它的最终目标是让机器能够像人一样具有分析学习能力，能够识别文字、图像和声音等数据。