

机器学习

维基百科，自由的百科全书

机器学习是近20多年兴起的一门多领域交叉学科，涉及概率论、统计学、逼近论、凸分析、计算复杂性理论等多门学科。机器学习理论主要是设计和分析一些让计算机可以自动“学习”的算法。机器学习算法是一类从数据中自动分析获得规律，并利用规律对未知数据进行预测的算法。因为学习算法中涉及了大量的统计学理论，机器学习与推断统计学联系尤为密切，也被称为统计学习理论。算法设计方面，机器学习理论关注可以实现的，行之有效的学习算法。很多推论问题属于无程序可循难度，所以部分的机器学习研究是开发容易处理的近似算法。

机器学习已广泛应用于数据挖掘、计算机视觉、自然语言处理、生物特征识别、搜索引擎、医学诊断、检测信用卡欺诈、证券市场分析、DNA序列测序、语音和手写识别、战略游戏和机器人等领域。

目录

- 1 定义
- 2 分类
- 3 算法
- 4 参考文献
 - 4.1 引用
 - 4.2 书目
- 5 外部链接
- 6 参见

定义

机器学习有下面几种定义：“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”。“机器学习是对能通过经验自动改进的计算机算法的研究”。“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。”一种经常引用的英文定义是：A computer program is said to learn from experience *E* with respect to some class of tasks *T* and performance measure *P*, if its performance at tasks in *T*, as measured by *P*, improves with experience *E*.

分类

机器学习可以分成下面几种类别：

- 监督学习从给定的训练数据集中学习出一个函数，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是包括输入和输出，也可以说是特征和目标。训练集中的目标是由人标注的。常见的监督学习算法包括回归分析和统计分类。
- 无监督学习与监督学习相比，训练集没有人为标注的结果。常见的无监督学习算法有聚类。
- 半监督学习介于监督学习与无监督学习之间。
- 增强学习通过观察来学习做成如何的动作。每个动作都会对环境有所影响，学习对象根据观察到的周围环境的反馈来做出判断。

算法

具体的机器学习算法有：

- 构造条件概率：回归分析和统计分类
 - 人工神经网络
 - 决策树
 - 高斯过程回归
 - 线性判别分析
 - 最近邻居法
 - 感知器
 - 径向基函数核
 - 支持向量机
- 通过再生模型构造概率密度函数：
 - 最大期望算法
 - graphical model：包括贝叶斯网和Markov随机场
 - Generative Topographic Mapping
- 近似推断技术：
 - 马尔可夫链
 - 蒙特卡罗方法
 - 变分法
- 最优化：大多数以上方法，直接或者间接使用最优化算法。

参考文献

引用

书目

- Bishop, C. M. (1995). 《模式识别神经网络》，牛津大学出版社。ISBN 0-19-853864-2
- Bishop, C. M. (2006). 《模式识别与机器学习》，Springer。ISBN 978-0-387-31073-2
- Richard O. Duda, Peter E. Hart, David G. Stork (2001) 《模式分类》（第2版），Wiley, New York, ISBN 0-471-05669-3.
- MacKay, D. J. C. (2003). 《信息理论、推理和学习算法》(<http://www.inference.phy.cam.ac.uk/mackay/itila/>)，剑桥大学出版社。ISBN 0-521-64298-1
- Mitchell, T. (1997). 《机器学习》，McGraw Hill. ISBN 0-07-042807-7
- Sholom Weiss, Casimir Kulikowski (1991). Computer Systems That Learn, Morgan Kaufmann. ISBN 1-55860-065-5

外部链接

- UCI description (<http://www.ics.uci.edu/~mlearn/Machine-Learning.html>)
- 机器学习软件Weka (<http://www.cs.waikato.ac.nz/~ml/>)
- Pablo Castro主页 (<http://www.dca.fee.unicamp.br/~pablo>)
- 机器学习网邮件列表 (<http://www.mlnet.org/>)
- 机器学习和自然语言处理-弗莱堡大学 (<http://www.informatik.uni-freiburg.de/~ml/>)
- 机器学习和数据挖掘, 生物信息学小组, 慕尼黑工业大学 (<http://www.kramer.in.tum.de/>)
- 机器学习和生物计算-Bristol大学 (<http://www.cs.bris.ac.uk/Research/MachineLearning/>)
- 机器学习和应用统计学@微软研究 (<http://research.microsoft.com/mlas/>)
- 机器学习研究月刊 (<http://jmlr.csail.mit.edu/>)
- 机器学习期刊 (<http://pages.stern.nyu.edu/~fprovost/MLJ/>)
- 机器学习-Kmining, 数据挖掘和KDD科学参考 (http://www.kmining.com/info_conferences.html)
- Book “智能系统社区 (<http://www.intelligent-systems.com.ar/intsys/index.htm>)” by Walter Fritz
- 开放目录项目 (http://dmoz.org/Computers/Artificial_Intelligence/Machine_Learning/)
- 机器学习论文-CiteSeer (<http://citeseer.ist.psu.edu/cis?q=machine+learning>)
- Orange, 使用Python脚本语言的机器学习组件和可视化编程接口 (<http://www.ailab.si/orange>)

参见

- 人工智能
- 计算智能
- 数据挖掘
- 模式识别
- 机器学习方面重要出版物（计算机科学）
- 机器学习方面重要出版物（统计学）
- 自主控制机器人
- 归纳逻辑编程
- 决策树
- 神经网络
- 强化学习
- 贝叶斯学习
- 最近邻居法
- 计算学习理论

取自“<https://zh.wikipedia.org/w/index.php?title=机器学习&oldid=35537660>”

- 本站的全部文字在知识共享 署名-相同方式共享 3.0协议之条款下提供，附加条款亦可能应用（请参阅使用条款）。

Wikipedia®和维基百科标志是维基媒体基金会的注册商标；维基™是维基媒体基金会的商标。

维基媒体基金会是在美国佛罗里达州登记的501(c)(3)免税、非营利、慈善机构。