

机器学习技术发展的综述与展望

孟子流¹, 李腾龙^{1,2}

(1. 河南省建筑科学研究院有限公司科研中心, 河南 450001;

2. 郑州大学, 信息工程学院, 河南 450001)

摘要: 基于机器学习技术的发展回顾, 阐述机器学习的技术类型, 包括决策树和基于逻辑的学习、神经网络的连接学习、统计学习支持向量机, 分析工作流程与分类。

关键词: 信息技术, 计算机工程, 机器学习, 神经网络。

中图分类号: TP18 文章编号: 1674-2583(2020)10-0056-02

DOI: 10.19339/j.issn.1674-2583.2020.10.024

中文引用格式: 孟子流, 李腾龙. 机器学习技术发展的综述与展望[J]. 集成电路应用, 2020, 37(10): 56-57.

Review and Prospect of Machine Learning Technology

MENG Ziliu¹, LI Tenglong^{1,2}

(1. Research Center of Henan Academy of Building Sciences Co., Ltd, Henan 450001, China.

2. School of information engineering, Zhengzhou University, Henan 450001, China.)

Abstract — Based on the review of the development of machine learning technology, this paper describes the types of machine learning technology, including decision tree and logic based learning, neural network connection learning, statistical learning support vector machine, and analyzes the workflow and classification.

Index Terms — information technology, computer engineering, machine learning, neural network.

0 引言

机器学习是利用计算机技术, 对于采集的数据做计算分析, 并通过不断的改进计算方法, 来提高完成特定系统的准确性。这个特定系统即为学习算法, 而机器学习则是将数据输入给学习算法, 算法根据输入的数据生成计算模型。用另一组测试数据输入计算模型, 该模型即可输出判断结果。机器学习就是完成以上学习算法的学科。

1 机器学习发展三个时期

人工智能从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放。1980年第一届机器学习研讨会IWML在美国卡耐基梅隆大学(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 and 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: ①R. S. Michalski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。

2 机器学习技术

(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

作者简介: 孟子流, 河南省建筑科学研究院有限公司科研中心, 研究方向: 机器学习与人工智能。

收稿日期: 2020-07-06, 修回日期: 2020-09-10。

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。（3）统计学习支持向量机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。

3 机器学习工作流程与分类

（1）机器学习工作流程，如图1所示：①数据获取与清洗：采集数据，并将数据进行清洗分类，区分为训练数据集，验证数据集，测试数据集。②构建模型：参考训练数据集构建合理的特征模型。③验证模型：在验证数据集上，对构建好的特征模型验证有效性，评估模型的可信度水平。④评估模型：在测试数据集上，对验证过的特征模型进行评估测试，并使用新数据对评估完成的特征模型做数据预测。⑤模型调优：采用参数调整，结果优化，调整权重，Bad-case分析等方法来提高算法的性能。

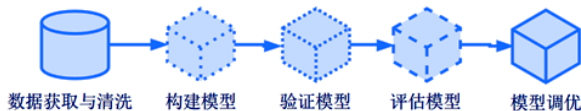


图1 机器学习工作流程图

（2）机器学习的分类。机器学习按照学习形式可以分为监督学习(Supervised learning)和非监督学习(Unsupervised learning)，还可以细分半监督学习和强化学习^[1-3]。①监督学习。监督学习(Supervised Learning)是指给定目标样本的值，机器学习算法通过计算得出目标样本和特征参数的关系。也就要求输入的数据是有标签的样本数据。分类树和回归树就是典型的

监督式学习方法。常见的监督式学习方法有很多，包括朴素贝叶斯分类器，支持向量机(SVM)，K最近邻算法(KNN, K-NearestNeighbor)，线性回归(LR, Linear Regreesion)，人工神经网络(ANN, Artificial Neural Network)，最小二乘法，决策树等。②非监督学习。非监督学习(Unsupervised Learning)也称无监督学习，与监督式学习不同，非监督学习所给定的数据集不需要添加标签，数据没有给定目标值和类别信息。这就要求根据数据直接建立模型，因此我们可以将数据降低到二维，三维，从而可以更好地在二维、三维面上展示数据信息。聚类和密度估计都是非监督学习的表示类型。③半监督学习。半监督学习(Semi-supervised Learning)是监督式学习和非监督学习相融合的一种学习方法。这类学习方法通常解决的问题是给定数据，有些是有标签的，有些是无标签的。比较常见的半监督学习方法有归纳学习法(Inductive learning)和传导推理法(Transductive Learning)。④强化学习。强化学习(Reinforcement Learning)又称为增强学习、评价学习，它是基于环境的学习方法，其特点是模型系统需要从环境中挖掘信号。环境给出少量信息后，模型系统需要通过自我调整，而使奖励函数(Rewrad function)的值最大。通俗一点讲就是根据给定的参数先做出行动，得到一个结果，通过判断结果的好坏来对之前的行为做一个分析评价。再调整算法从而得到更好的结果。强化学习的重点和难点在于如何设计奖励机制(Reward mechanisms)，强化学习目前只要应用于机器人控制领域。

机器学习的核心就是数据清洗、算法模型、计算机运算能力。我们研究的核心就是算法，也就是针对某一应用需求场景，选择各种合适的函数公式进行构建和叠加形成一个处理数据的模型（也就是常说的训练数据模型）。模型的作用就是能够将普通数据输入进去，经过处理后，输出数据或者一个结果，供我们参考或者使用的。

4 结语

本文介绍了从1950年代至今，机器学习技术的发展以及发展过程中存在的问题，总结并详细描述了机器学习技术的发展和分类。对机器学习技术做了归纳总结，帮助认识机器学习技术的发展史，对后续学习研究机器学习技术与算法有一定参考意义。

参考文献

- [1] 李旭然, 丁晓红. 机器学习的五大类别及其主要算法综述[J]. 软件导刊, 2019, 18(07): 4-9.
- [2] 刘凯, 张立民, 周立军. 深度学习在信息推荐系统的应用综述[J]. 小型微型计算机系统, 2019, 40(04): 738-743.
- [3] 黄亚强. 以机器学习为基础的人工智能[J]. 电子技术与软件工程, 2018(08): 257.