Near笔记之机器学习概述

问题是打开未知世界的钥匙

1. 如何定义“机器学习”这个概念?
2. 机器学习是为了解决什么实际问题而提出?
3. 机器学习的原理和理论基础是什么?
4. 了解机器学习的发展历史，可以预见未来发展的趋势吗?
5. 机器学习有哪些常用方法?
6. 机器学习方法的评价标准和选用标准有哪些?

什么是“机器学习”

> 如何定义“机器学习”这个概念?
>
> ***概念定义***
>
> 机器学习是一门多领域交叉学科，涉及概率论、统计学、逼近论、凸分析、算法复杂度理论等多门学科。专门研究计算机怎样模拟或实现人类的学习行为，以获取新的知识或技能，重新组织已有的知识结构使之不断改善自身的性能。 在算法的指导下，通过对大量的数据进行训练，分析出其中隐藏的结构或规律，并不断提升算法性能，对事件的发生进行判断或预测。其目标是让计算机学会如何模拟人进行思维和创造，不断的学习新的知识和技能。
>
> ***相关定义***

1. “最基本的机器学习是使用算法解析数据，从中学习，然后对世界上某事做出决定或预测的做法。” - Nvidia
2. “机器学习是让计算机在没有明确编程的情况下采取行动的科学。” - 斯坦福
3. “机器学习基于可以从数据中学习而不依赖于基于规则的编程的算法。” - 麦肯锡公司
4. “机器学习算法可以通过推广实例来弄清楚如何执行重要任务。” - 华盛顿大学
5. “机器学习领域旨在回答这样一个问题：”我们如何建立能够根据经验自动改进的计算机系统，以及管理所有学习过程的基本法则是什么？“ - 卡内基梅隆大学

小结

机器学习是一门多领域交叉学科，主要研究计算机怎样模拟或实现人类的学习行为，以不断完善自身算法来提供更好的服务，特别是如何在经验（数据）学习中改善具体算法的性能。

“机器为何学习”

> 机器学习是为了解决什么实际问题而提出？

计算机解决具体问题的传统过程，一般是根据已经编辑好的程序指令按序处理得到确定结果的过程。这个过程要求有明确的算法流程——我们人类可以用公式或逻辑来表示的流程。而现实生活中，有很多问题并不能找到（或目前条件还不能找到）一个明确的算法流程或公式逻辑，但是，因为有以往的数据记录，这时候就要求有一种方法能够从过去的经验数据中发现规律或模式，以便用于形成一个拟合程度高的近似算法来解决实际问题。

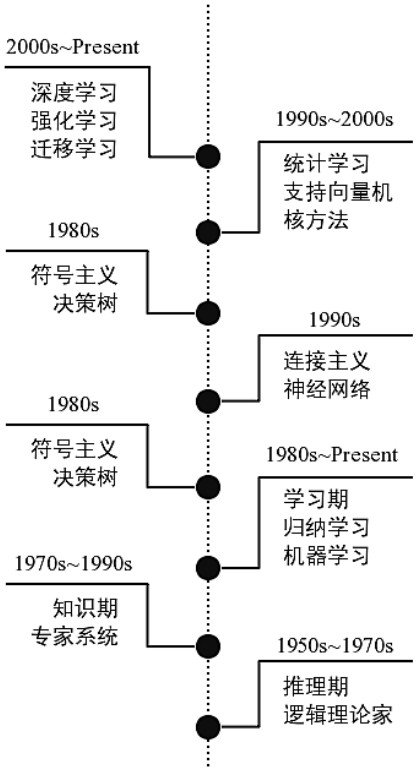
小结

机器学习是为了解决那些有现实数据但目前还未能用具体逻辑流程来表示其问题解决流程的复杂问题。期望通过其过往数据，发现其规律或者找到一个拟合程度高的算法流程来解决实际问题。

机器学习的发展

> 从机器学习的历史发展过程，能了解到什么？

历史简图



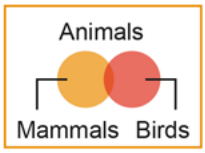
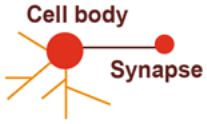
学派分类

A look at
Machine learning evolution

Overview

For decades, individual "tribes" of artificial intelligence researchers have vied with one another for dominance. Is the time ripe now for tribes to collaborate? They may be forced to, as collaboration and algorithm blending are the only ways to reach true artificial general intelligence (AGI). Here's a look back at how machine learning methods have evolved and what the future may look like.

What are the five tribes?

Symbolists	Bayesians	Connectionists	Evolutionaries	Analogizers
				
Use symbols, rules, and logic to represent knowledge and draw logical inference	Assess the likelihood of occurrence for probabilistic inference	Recognize and generalize patterns dynamically with matrices of probabilistic, weighted neurons	Generate variations and then assess the fitness of each for a given purpose	Optimize a function in light of constraints ("going as high as you can while staying on the road")
Favored algorithm Rules and decision trees	**Favored algorithm** Naive Bayes or Markov	**Favored algorithm** Neural networks	**Favored algorithm** Genetic programs	**Favored algorithm** Support vectors

Source: Pedro Domingos, *The Master Algorithm*, 2015

1. ***符号学派（Symbolists）***：是使用基于规则的符号系统做推理的人。爱用方法：规则和决策树。
2. ***贝叶斯学派（Bayesians）***：是使用概率规则及其依赖关系进行推理的一派。爱用方法：朴素贝叶斯或马尔科夫。
3. ***联结学派（ConnecTIonists）***：这一派的研究者相信智能起源于高度互联的简单机制。最新的形式是深度学习。爱用方法：神经网络。
4. ***进化学派（EvoluTIonists）***：是应用进化的过程，例如交叉和突变以达到一种初期的智能行为的一派。爱用方法：遗传算法。
5. ***类推学派（The analogizers）***：更多地关注心理学和数学最优化，通过外推来进行相似性判断。爱用方法：支持向量机（SVM）。

Phases of evolution

1980s

Predominant tribe
Symbolists

Architecture
Server or mainframe

Predominant theory
Knowledge engineering

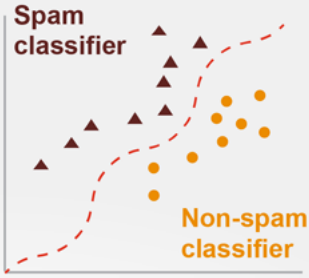
Basic decision logic:
Decision support systems with limited utility

1990s to 2000

Predominant tribe
Bayesians

Architecture
Small server clusters

Predominant theory
Probability theory



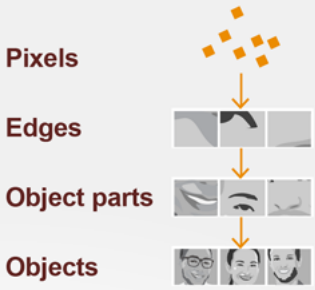
Classification:
Scalable comparison and contrast that's good enough for many purposes

Early to mid-2010s

Predominant tribe
Connectionists

Architecture
Large server farms (the cloud)

Predominant theory
Neuroscience and probability



Recognition:
More precise image and voice recognition, translation, sentiment analysis, etc.

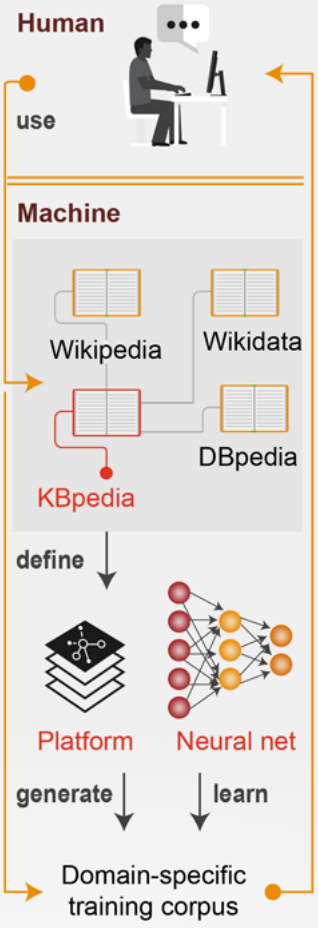
The tribes see fit to collaborate and blend their methods

Late 2010s

Predominant tribe
Connectionists + Symbolists

Architecture
Multiple clouds

Predominant theory
Memory neural networks, large-scale integration, and reasoning over knowledge

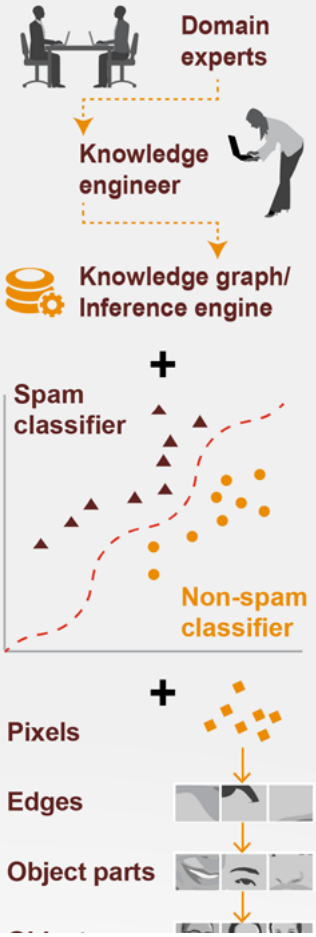


2020s+

Predominant tribe
Connectionists + Symbolists + Bayesians + ...

Architecture
Clouds and fog

Predominant theory
Networks when sensing, but rules when reasoning and acting

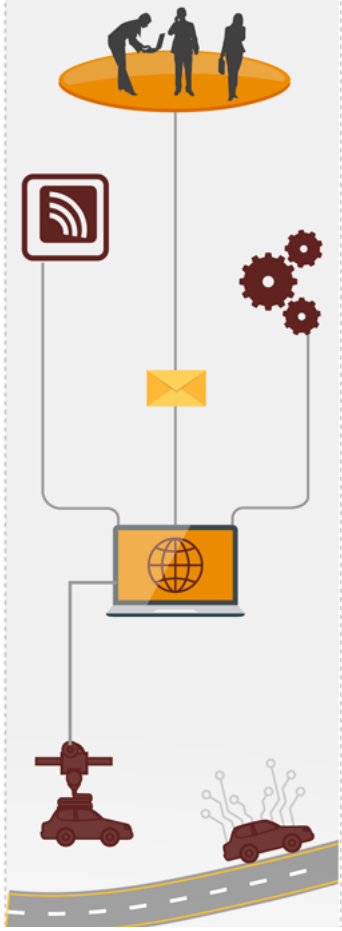


2040s+

Predominant tribe
Algorithmic convergence

Architecture
Server ubiquity

Predominant theory
Best-of-breed meta-learning



机器学习的原理

> 机器学习的原理和理论基础是什么?

机器学习是计算机基于数据构建概率统计模型并运用模型对数据进行预测与分析的学科。 可说是从数据中来，到数据中去。因此从大量现象中提取反复出现的规律与模式是可能的。 ***机器学习 = 数据（data） + 模型（model） + 优化方法（optimal strategy）***

数据

在机器学习中，数据并非通常意义上的数量值，而是对于对象某些性质的描述。被描述的性质叫作属性，属性的取值称为属性值，不同的属性值有序排列得到的向量就是数据，也叫实例。根据线性代数的知识，数据的不同属性之间可以视为相互独立，因而每个属性都代表了一个不同的维度，这些维度共同张成了特征空间。每一组属性值的集合都是这个空间中的一个点，因而每个实例都可以视为特征空间中的一个向量，即特征向量。通常来说，数据都是从收集到输入算法模型，需要经过收集，清理，拆分（分为训练模型用的训练集和评估模型的测试集）等过程。

模型

激活函数

线性模型的表达能力不够，激活函数可以增加神经网络模型的非线性，提升神经网络模型表达能力（数据往往线性不可分）。

1. sigmoid函数
2. tanh函数
3. ReLU函数
4. Leaky ReLU函数（PReLU）
5. ELU函数
6. MaxOut函数
7. softmax函数

损失函数

损失函数是用来估量你模型的预测值f(x)与真实值Y的不一致程度，它是一个非负实值函数,通常使用L(Y, f(x))来表示，损失函数越小，模型的鲁棒性就越好。

1. 0-1损失函数

$$L(y, f(x)) = \begin{cases} 0, & \text{y = f(x);} \\ 1, & \text{y} \neq \text{f(x);} \end{cases}$$

2. 绝对值损失函数

$$L(y, f(x)) = |y - f(x)|$$

3. 平方损失函数

$$L(y, f(x)) = (y - f(x))^2$$

4. log对数损失函数

$$L(y, f(x)) = log(1 + e^{-yf(x)})$$

5. 指数损失函数

$$L(y, f(x)) = exp(-yf(x))$$

6. Hinge损失函数

$$L(w, b) = max0, 1 - yf(x)$$

优化方法

梯度下降是最常用的优化方法之一，它使用梯度的反方向$\nabla_\theta J(\theta)$更新参数θ，使得目标函数$J(\theta)$达到最小化的一种优化方法，这种方法我们叫做梯度更新.

(全量)梯度下降

$$\theta = \theta - \eta \nabla_\theta J(\theta)$$

随机梯度下降

$$\theta = \theta - \eta \nabla_\theta J(\theta; x^{(i)}, y^{(i)})$$

小批量梯度下降

$$\theta = \theta - \eta \nabla_\theta J(\theta; x^{(i:i+n)}, y^{(i:i+n)})$$

引入动量的梯度下降

$$\begin{cases} v_t = \gamma v_{t-1} + \eta \nabla_\theta J(\theta) \\ \theta = \theta - v_t \end{cases}$$

自适应学习率的Adagrad算法

$$\begin{cases} g_t = \nabla_\theta J(\theta) \\ \theta_{t+1} = \theta_{t,i} - \frac{\eta}{\sqrt{G_t + \varepsilon}} \cdot g_t \end{cases}$$

牛顿法

$$\theta_{t+1} = \theta_t - H^{-1} \nabla_\theta J(\theta_t)$$

其中: t: 迭代的轮数 η: 学习率

G_t: 前t次迭代的梯度和

ε :很小的数,防止除0错误

H: 损失函数相当于θ的Hession矩阵在θ_t处的估计

机器学习方法分类

> 机器学习有哪些常用方法?

按学习方式分类

有监督、无监督、半监督、强化学习

> P.S. 后续详细了解再扩展

按任务类型分类

回归、分类、聚类、降维 生成模型与判别模型

评价标准和选用标准

> 机器学习方法的评价标准和选用标准有哪些?

模型评估指标:

1. MSE(Mean Squared Error)
2. MAE(Mean Absolute Error)
3. RMSE(Root Mean Squard Error)
4. Top-k准确率
5. 混淆矩阵——衡量的是一个分类器分类的准确程度。

复杂度度量：偏差与方差、过拟合与欠拟合、结构风险与经验风险、泛化能力、正则化

模型选择

正则化、交叉验证

采样：样本不均衡

特征处理：归一化、标准化、离散化、one-hot编码

模型调优

网格搜索寻优、随机搜索寻优、贝叶斯优化算法

> P.S. 先放放。。。

参考文献

1. https://github.com/datawhalechina/team-learning/blob/master/%E5%88%9D%E7%BA%A7%E7%AE%97%E6%B3%95%E6%A2%B3%E7%90%86 (https://github.com/datawhalechina/team-learning/blob/master/%E5%88%9D%E7%BA%A7%E7%AE%97%E6%B3%95%E6%A2%B3%E7%90%86)
2. https://github.com/sldyns/Introduce_to_Machine_Learning (https://github.com/sldyns/Introduce_to_Machine_Learning)
3. https://www.visualistan.com/2017/12/a-look-at-machine-learning-evolution.html (https://www.visualistan.com/2017/12/a-look-at-machine-learning-evolution.html)
4. http://www.elecfans.com/rengongzhineng/669273.html (http://www.elecfans.com/rengongzhineng/669273.html)
5. https://blog.csdn.net/yoggieCDA/article/details/101465011 (https://blog.csdn.net/yoggieCDA/article/details/101465011)