

PaddlePaddle李宏毅机器学习特训营笔记——机器学习概述

原创

[从零开始机器学习](#) 于 2021-04-14 23:34:17 发布 182 收藏 2

文章标签: [人工智能](#) [机器学习](#) [大数据](#) [算法](#)

版权声明: 本文为博主原创文章, 遵循 [CC 4.0 BY-SA](#) 版权协议, 转载请附上原文出处链接和本声明。

本文链接: https://blog.csdn.net/weixin_47291626/article/details/115709796

版权

本文的目录如下:

1. 机器学习引入
2. 什么是机器学习?
3. 机器学习的分类
4. 机器学习应用场景
5. 课程信息

1. 机器学习引入

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。

机器学习是人工智能的一个子集, 它是人工智能的核心, 是使计算机具有智能的根本途径。它表示计算机使用算法从数据中学习, 允许机器识别不同模式, 而组织可以通过各种方式运用此技术。专家表示, 机器学习可以帮助组织通过非同以往的规模和范围执行任务。因此, 它可以加快工作速度, 减少错误, 提高准确性, 从而对员工和客户有所帮助。此外, 以创新为导向的组织正在寻找利用机器学习的机会, 不仅提高效率, 还能激发新的商业机会, 使公司和业务在市场中脱颖而出。Addressable Markets首席策略师表示:“机器学习正在改善几乎所有功能和过程的自动化, 因为它能根据不断变化的条件进行适应和操作。

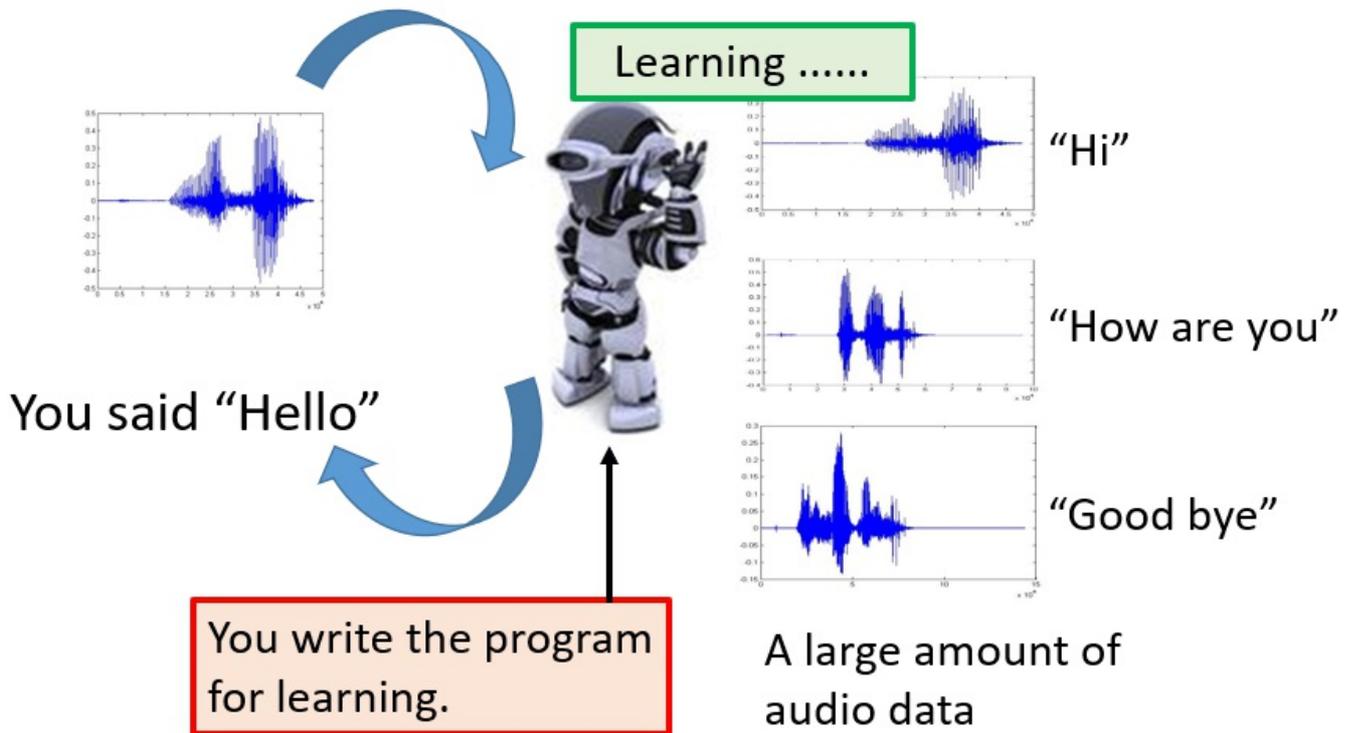
在《The Master Algorithm》这本书中写到: 你可能不知道, 但是机器学习早就已经环绕在你身边了。当你在搜索引擎搜索的时候, 机器学习来帮助搜索引擎判断哪个结果更适合你 (也在判断哪个广告更适合你)。大部分垃圾邮件不会被你注意到, 是因为他们早就被机器学习过滤掉了。你去亚马逊买书或者去迅雷视频看外文电影时, 是机器学习在推荐书目并且帮你匹配字幕。知乎用机器学习来决定哪个回答更适合排在前面 (我不确定, 希望知乎的朋友说下), 微博也在你的时间线上做了同样的事情。可以说, 你在用电脑的时候总是会在某处碰到机器学习。传统来说让电脑去做什么事情的唯一方法就是去写下一个算法解释如何实现。但机器学习算法并不是这样, 它们会自己找到答案。换句话说, 在机器学习中, 并不是我们给机器编程, 而是机器自己给自己编程。

我们可能还没有注意到, 我们的生活已经离不开机器学习了。因此, 深入了解机器学习显得十分有必要。

2. 什么是机器学习?

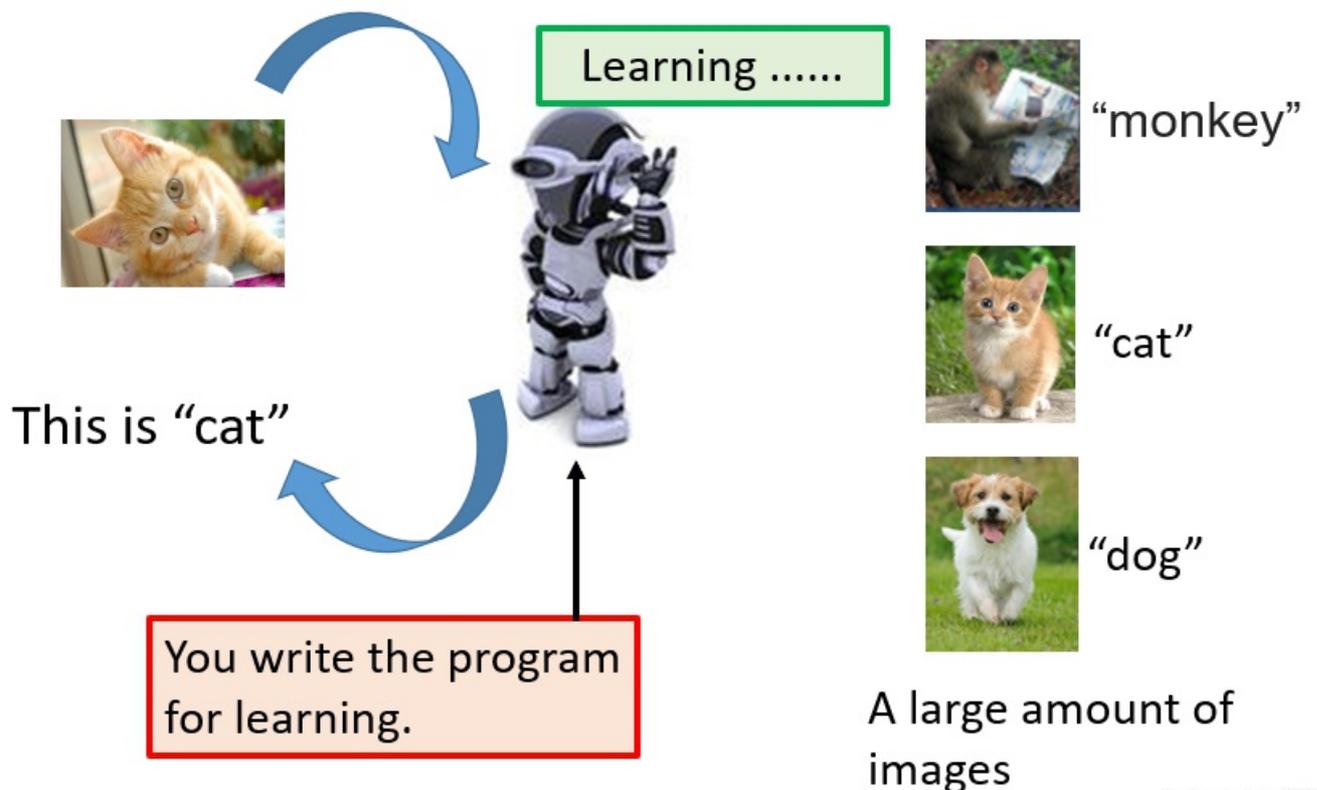
那究竟何谓机器学习呢？在给概念之前先来看李宏毅老师PPT中的几张图，我们就能对机器学习产生初步的认识了。

What is Machine Learning?



上图是一个语音处理的例子，机器通过输入的语音信号，输出说的语句。

What is Machine Learning?



上图是一个图像分类的例子，机器根据输入的图片，判断并输出其所属的类别。

根据上面两张图，我们不难看出，所谓机器学习就是通过算法进行学习，然后模拟人的智能进行一系列决策的过程，称为机器学习。

机器学习的本质就是寻找一个拟合函数，然后给定输入，产生相应的输出。

Machine Learning ≈ Looking for a Function

- Speech Recognition

$$f(\text{audio waveform}) = \text{"How are you"}$$

- Image Recognition

$$f(\text{cat image}) = \text{"Cat"}$$

- Playing Go

$$f(\text{Go board}) = \text{"5-5"} \quad (\text{next move})$$

- Dialogue System

$$f(\text{"Hi"} \text{ (what the user said)}) = \text{"Hello"} \text{ (system response)}$$

学习的过程就是通过优化寻找一个最优函数的过程。

优化的算法很多，最常见也是最基础的算法是梯度下降算法，梯度下降法是最早接触的优化算法，也是应用最广泛的优化算法。

Image Recognition:

Framework

$$f(\text{cat image}) = \text{"cat"}$$



$$f_1(\text{cat image}) = \text{"cat"} \quad f_2(\text{cat image}) = \text{"money"}$$



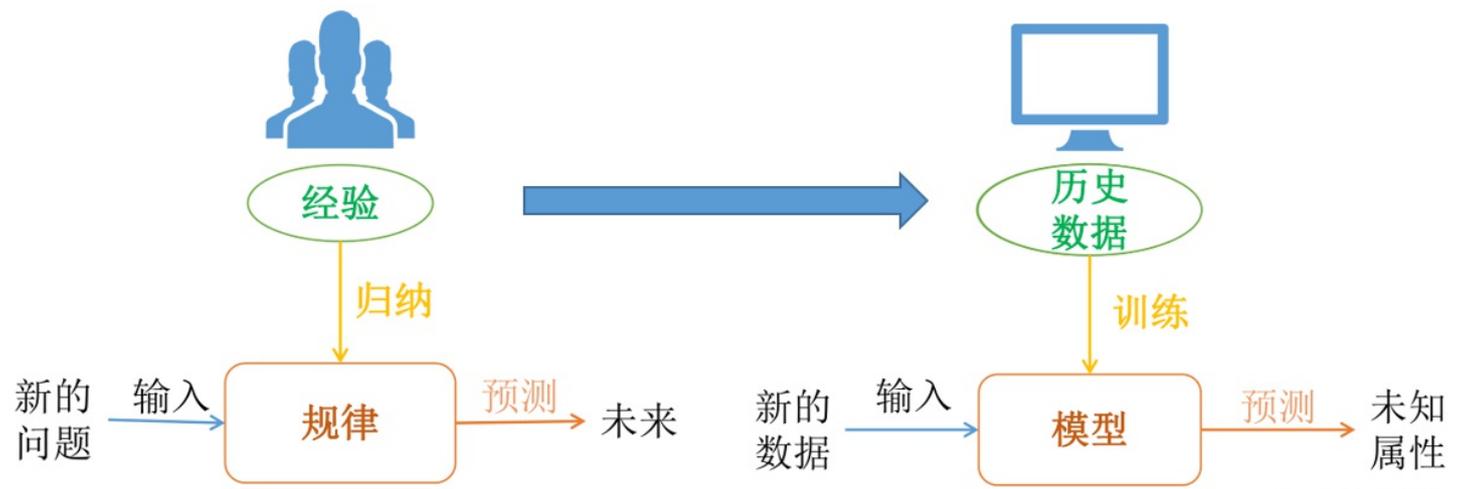
https://img-prod-netease.com_3729628

如何寻找最优的函数呢？这里不得不提一下机器学习的经典定义：

对于机器来说，**经验**是通过**数据**传达的。机器学习的主要研究内容就是从数据中产生**模型**的算法，也即**学习算法**。Mitchell 给出一个更为形式化的定义，假设：

- **P**：计算机程序在某任务类**T**上的**性能**。
- **T**：计算机程序希望实现的**任务**。
- **E**：表示**经验**，即**历史的数据集**。

若该计算机程序通过利用经验 **E** 在任务 **T** 上获得了性能 **P** 的改善，则称该程序对**E**进行了学习。

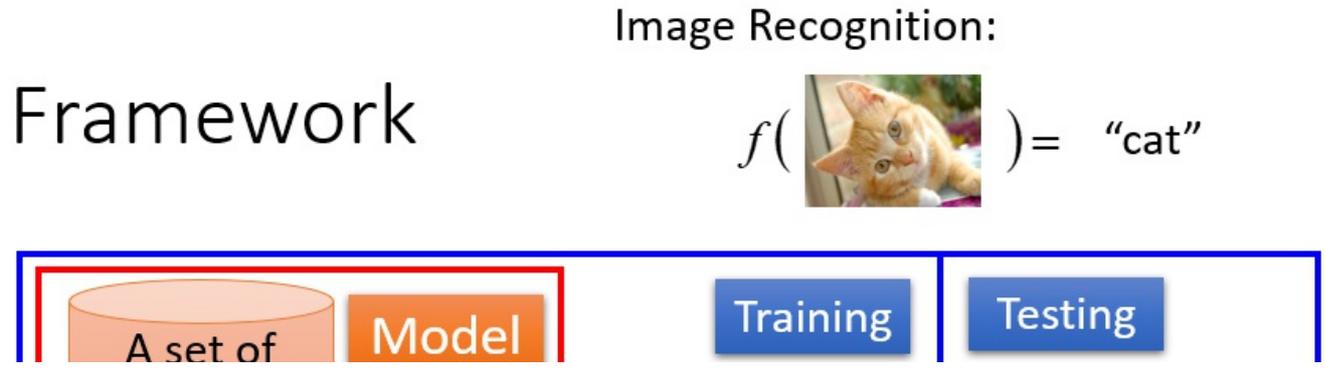


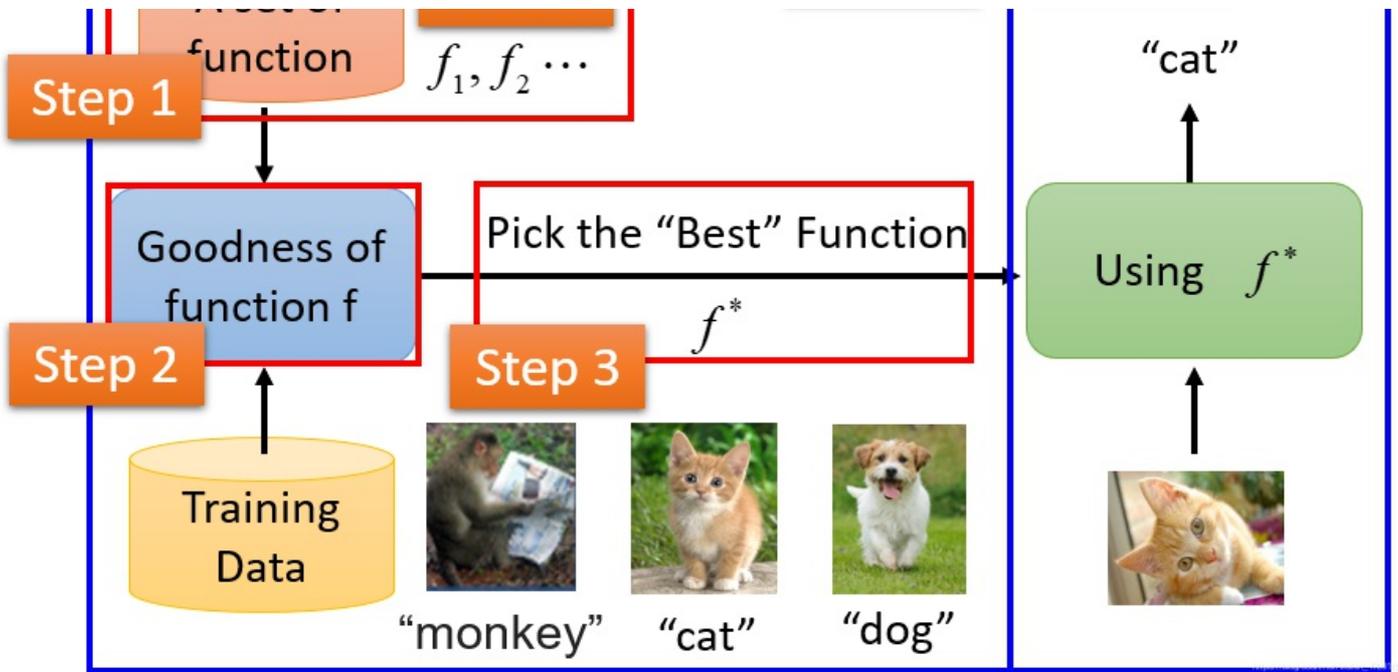
https://img-prod-netease.com_42291028

正如我们根据过去的经验来判断明天的天气，吃货们希望从购买经验中挑选一个好瓜，那能不能让计算机帮助人类来实现这个呢？机器学习正是这样的一门学科，人的“经验”对应计算机中的“数据”，让计算机来学习这些经验数据，生成一个算法模型，在面对新的情况中，计算机便能作出有效的判断，这便是**机器学习**。

要进行机器学习，首先要有数据，我们可以收集一组结构相同的记录，这组记录的集合就称为**数据集**。但是也不是所有的机器学习都需要数据，下面我们会讲到。

机器学习的全流程如下所示：



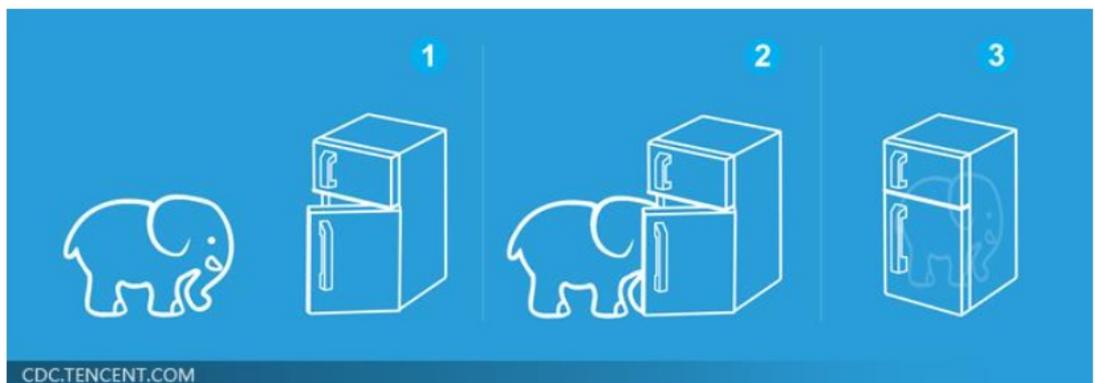


此处我们可以通过李宏毅教授的比喻来进行通俗的解释，机器学习就像把大象放进冰箱一样简单（bushi

Machine Learning is so simple



就好像把大象放进冰箱



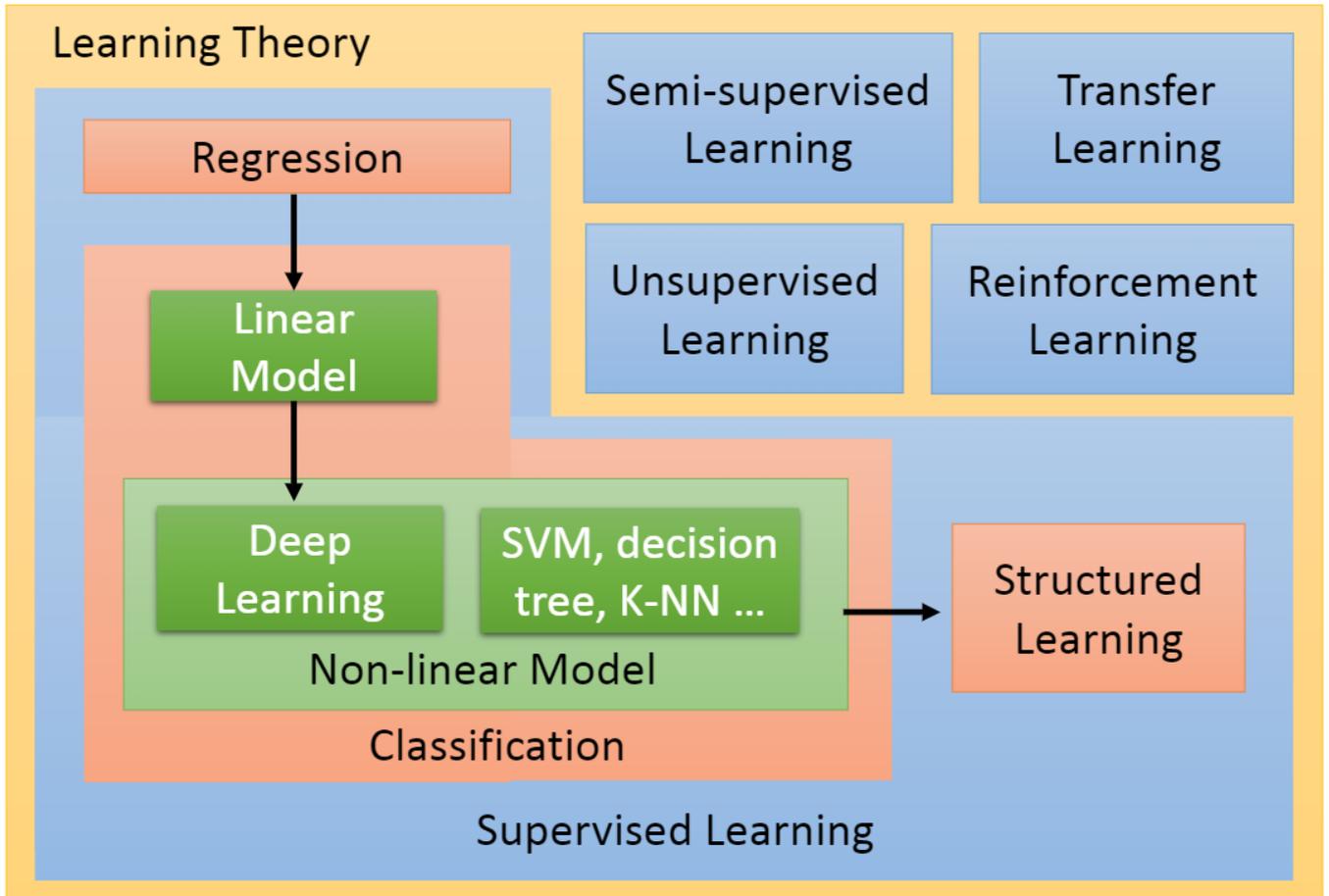
觉得文字太枯燥，可以点击这个机器学习科普进行更为直观的理解。

3. 机器学习的分类

机器学习是一个很宽泛的概念，有许多不同的分类方法，在本课程中主要按照下图进行分类。

Learning Map

■ scenario ■ task ■ method



我们根据学习的方式，也可以对机器学习进行如下分类。

1. 监督式学习

在监督式学习下，输入数据被称为“训练数据”，每组训练数据有一个明确的标识或结果，如对防垃圾邮件系统中“垃圾邮件”“非垃圾邮件”，对手写数字识别中的“1”，“2”，“3”等。在建立预测模型的时候，监督式学习建立一个学习过程，将预测结果与“训练数据”的实际结果进行比较，不断的调整预测模型，直到模型的预测结果达到一个预期的准确率。监督式学习的常见应用场景如分类问题和回归问题。常见算法有逻辑回归（Logistic Regression）和反向传递神经网络（Back Propagation Neural Network）。

2. 非监督式学习

在非监督式学习中，数据并不被特别标识，学习模型是为了推断出数据的一些内在结构。常见的应用场景包括关联规则的学习以及聚类等。常见算法包括Apriori算法以及k-Means算法。简言之，非监督式学习不需要数据即可进行学习。

3. 半监督式学习

顾名思义，这是一种介于监督学习和非监督式学习之间的一种学习方式。在此学习方式下，输入数据部分被标识，部分没有被标识，这种学习模型可以用来进行预测，但是模型首先需要学习数据的内在结构以便合理的组织数据来进行预测。应用场景包括分类和回归，算法包括一些对常用监督式学习算法的延伸，这些算法首先试图对未标识数据进行建模，在此基础上再对标识的数据进行预测。如图论推理算法（Graph Inference）或者拉普拉斯支持向量机（Laplacian SVM）等。

4. 强化学习

在这种学习模式下，输入数据作为对模型的反馈，不像监督模型那样，输入数据仅仅是作为一个检查模型对错的方式，在强化学习下，输入数据直接反馈到模型，模型必须对此立刻作出调整。常见的应用场景包括动态系统以及机器人控制等。常见算法包括Q-Learning以及时间差学习（Temporal difference learning）。AlphaGo战胜“石神”李世石，强化学习一战封神，被认为是最接近人类学习过程的方式，并使得强化学习走进人们的视野，使得人工智能开始闯进大众的视野。

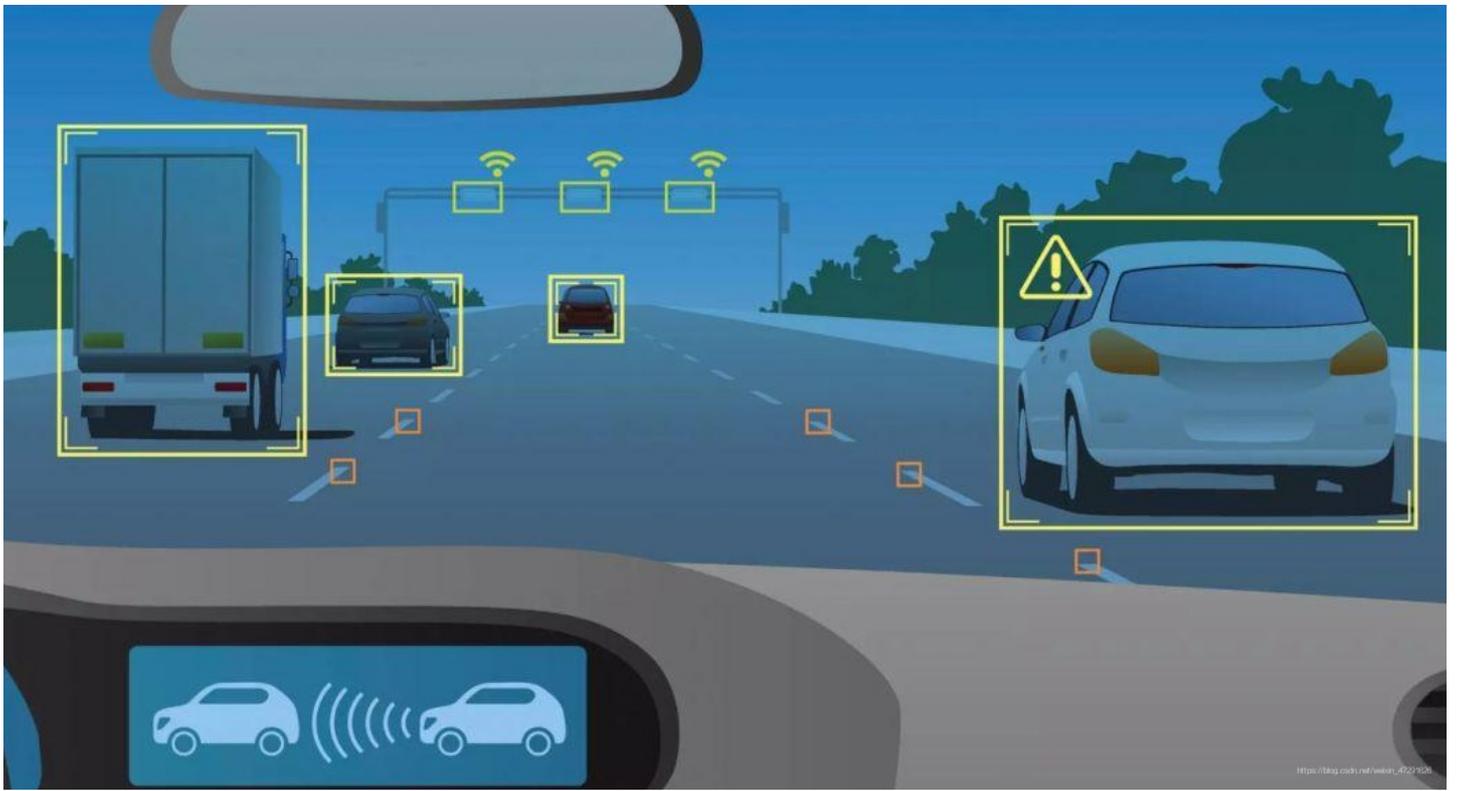


在企业数据应用的场景下，人们最常用的可能就是监督式学习和非监督式学习的模型。在图像识别等领域，由于存在大量的非标识的数据和少量的可标识数据，目前半监督式学习是一个很热的话题。而强化学习更多的应用在机器人控制及其他需要进行系统控制的领域。

[感兴趣的读者还可以点击机器学习分类](#)进行更为深入的了解。

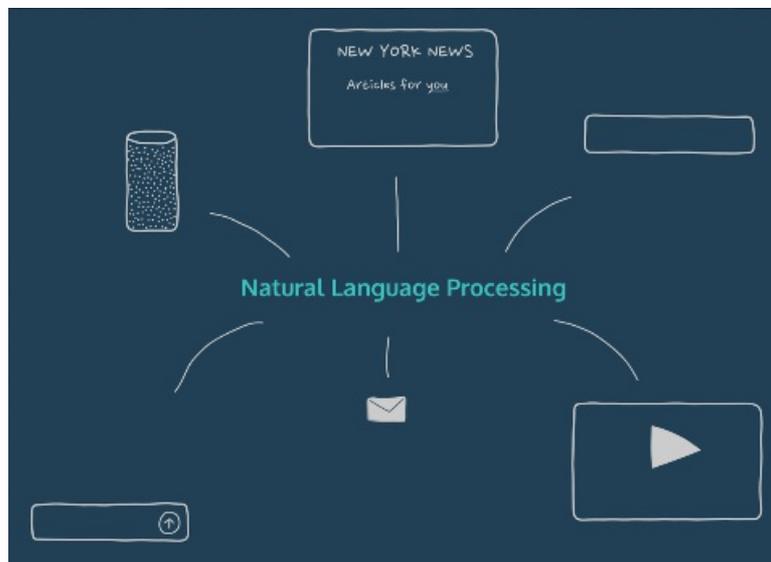
4. 机器学习应用场景

机器学习的应用场景非常多，目前落地场景最为广泛的是计算机视觉领域、自然语言处理和推荐系统。

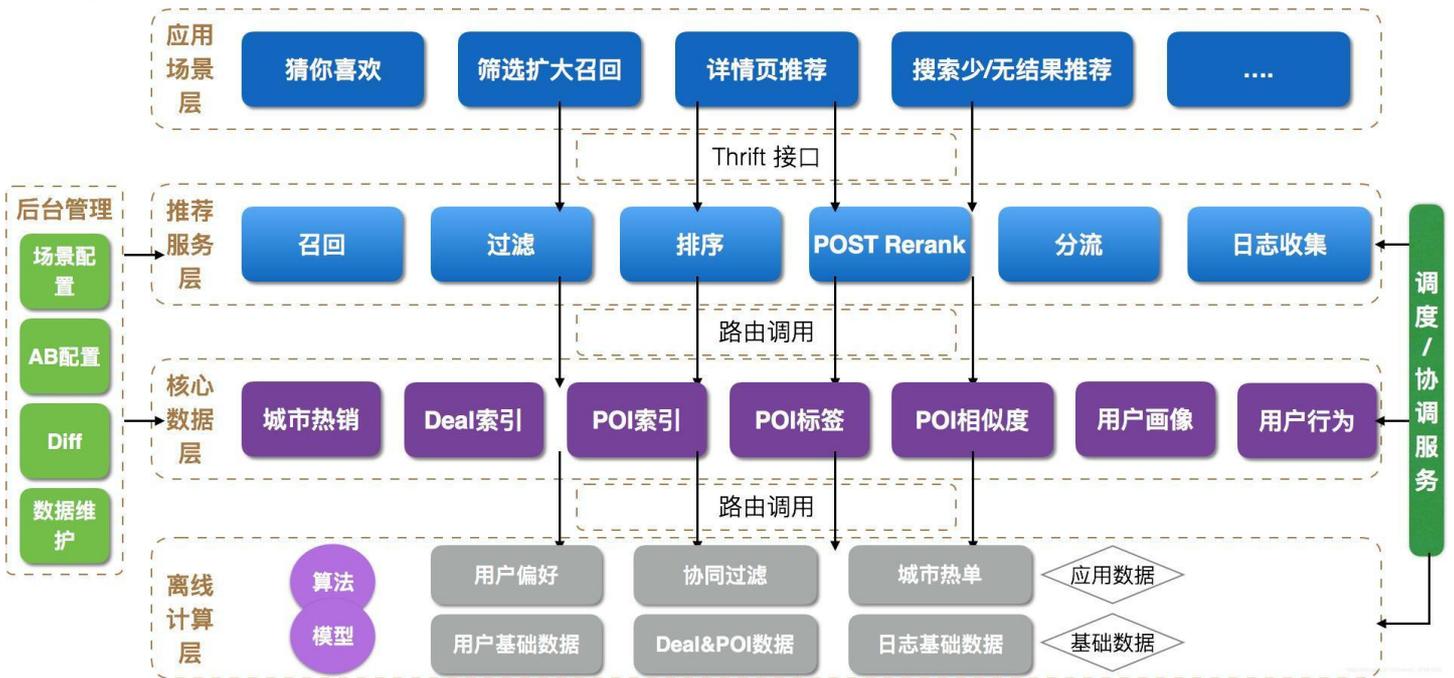


计算机视觉（Computational Vision）是由相机拍摄图像，通过电脑对图像中的目标进行识别和检测。可以说是机器学习在视觉领域的应用，是人工智能领域的一个重要部分。它的研究内容可以概括为：通过采集图片或视频，对图片或视频进行处理分析，从中获取相对应的信息。换言之就是运用照相机和计算机来获取我们所需的信息。这是机器学习目前应用最成功的领域，上图是计算机视觉在自动驾驶领域的一个例子。

它是一门包含领域很广的综合性学科。从现阶段的研究来看，计算机视觉试图建立一种人工系统，提出的越来越多的理论和技术的目的是为了从图像或者多维数据中获取信息。



自然语言处理主要处理的是文本类的问题，主要落地领域为机器翻译，文本分类与摘要，智能输入法，聊天机器人等。如下图即是利用基于循环神经网络的seq2seq或其他算法，利用大量语料训练的一个自动对联系统，尽管在某些语境上与人类还有很大差距，但是作为一台机器能做到这些，已经很酷了，不是吗？这个领域不像计算机视觉落地那样多，落地应用很少，主要还是这个领域很难。



推荐系统也是机器学习成功应用的一个场景。随着计算机领域技术的高速发展，电子商务时代的普及，个性化的推荐系统深入生活应用的各个方面。个性化推荐算法是推荐系统中最核心的技术，在很大程度上决定了电子商务推荐系统性能的优劣。除此之外，还有视频和新闻的个性化推荐，我们有时候会沉迷于刷视频和刷新闻的快感中，这是因为网站通过推荐系统推断出我们的喜好，然后不断推荐我们感兴趣的内容，我们会沉迷其中也就不难理解了。推荐系统应用也较广泛，可用于适合零售、媒体与娱乐、旅游与服务行业、教育、金融服务、政府、医疗保健、软件与互联网。

除此之外，机器学习深入我们生活的方方面面，说涉及到我们的衣食住行一点也不夸张。

5.课程信息

本文介绍了机器学习的概念、分类及其应用，有利于读者对机器学习领域建立一个初步的认识，有利于后续深入研究机器学习乃至人工智能领域。

想要更加深一步了解机器学习的知识，那就加入百度AI Studio的李宏毅机器学习特训营课程进行学习吧！

全部的目录如下：

项目	数据集	课程	比赛	认证	更多
 <p>李宏毅课程-机器学习</p> <p>飞桨深度学习学院</p> <p>2020/11/13 - 2025/12/31</p> <p>分享</p>		<p>课节 共18课节</p> <hr/> <p>课节1: 《机器学习》特训营开营仪式</p> <hr/> <p>课节2: 《机器学习》特训营实践研讨课(一)</p> <hr/> <p>课节3: 《机器学习》特训营实践研讨课(二)</p>			

📅 课节

📁 作业

1

📄 考试

🏆 比赛

1

讨论区

📢 公告栏

📖 课程介绍

👤 教师介绍

课节4: 《机器学习》特训营实践研讨课(三)

课节5: 《机器学习》特训营实践研讨课(四)

课节6: 机器学习介绍

课节7: 回归

课节8: 梯度下降

课节9: 分类

课节10: 支持向量机

课节11: 集成学习

课节12: 深度学习预备

课节13: 卷积神经网络

课节14: 循环神经网络

课节15: 半监督学习

课节16: 无监督学习

课节17: 深度生成模型

课节18: 迁移学习

https://blog.csdn.net/weixin_47291626

- 厲害的 AI ， AI 訓練師功不可沒
- 讓我們一起朝 AI 訓練師之路邁進





提到李宏毅老师，熟悉机器学习的读者应该都知道，他的机器学习视频被奉为领域内的经典视频之一，对于想要入门机器学习又想看中文讲解的人来说绝对是非常推荐的。他的课程中大量使用二次元来作为机器学习的例子，授课风格生动有趣，这样的宝藏老师的课程，你确定你不来？

让我们开始探索之旅,一起学习《李宏毅机器学习特训营》，开始宝可梦训练师成长之路吧！

下面附上本课程的传送门：<https://aistudio.baidu.com/aistudio/education/group/info/1978>

声明：本文章部分素材来自网络。



[创作打卡挑战赛](#) >

[赢取流量/现金/CSDN周边激励大奖](#)