

# 机器学习综述

赵晨阳

(德州市第一中学, 德州 253000)

**摘要:**本文介绍了机器学习的发展历程, 分析了机器学习的常见方法, 列举了机器学习在现代信息生活中的一些具体应用。

**关键词:**人工智能; 机器学习; 强化学习; 推荐系统

**doi:** 10.3969/J.ISSN.1672-7274.2018.01.081

**中图分类号:** TP18

**文献标识码:** A

**文章编码:** 1672-7274 (2018) 01-0109-02

## 1 引言

人工智能的主要三大驱动力: 大数据、机器学习、硬件 GPU, 本文主要从人工智能的算法——机器学习方面来展开论述。机器学习 (Machine Learning) 的本质是基于互联网的海量数据以及系统强大的并行运算能力, 让机器自主模拟人类学习的过程, 通过不断“学习”数据来做出智能决策行为。

## 2 机器学习的发展历程与相关要素

### 2.1 机器学习的发展历程

机器学习属于人工智能中一个较为年轻的分支, 可以大致分为以下三个发展历程:

第一阶段: 20 世纪 50 年代中期—60 年代中期, 这一时期处于萌芽时期。人们试图通过软件编程来操控计算机完成一系列的逻辑推理功能, 进而使计算机具有一定程度上类似人类一样的智能思考能力。然而这时期计算机所推理的结果远远没有达到人们对机器学习的期望。通过进一步研究发现, 只具有逻辑推理能力并不能使机器智能。研究者们认为, 使得机器拥有人工智能的前提, 还必须是拥有大量的先验知识<sup>[1]</sup>。

第二阶段: 20 世纪 60 年代中期—80 年代中期, 这一时期处于发展时期。人们试图利用自身思维提取出来的规则来教会计算机执行决策行为, 主流之力便是各式各样的“专家系统”, 然而这些系统总会面临“知识稀疏”的问题, 即面对无穷无尽的知识与信息, 人们无法总结出万无一失的规律。因此, 让机器自主学习的设想自然地浮出水面。基于 20 世纪 50 年代对于神经网络的研究, 人们开始研究如何让机器自主学习。

第三阶段: 20 世纪 80 年代—至今, 机器学习达到了一个繁荣时期。由于这一时期互联网大数据以及硬件 GPU 的出现, 使得机器学习脱离了瓶颈期。机器学习开始爆炸式发展, 开始成为了一门独立热门学科并且被应用到各个领域。各种机器学习算法不断涌现, 而利用深层次神经网络的深度学习也得到进一步发展。同时, 机器学习的蓬勃发展还促进了其他分支的出现, 例如模式识别, 数据挖掘, 生物信息学和自动驾驶等等。

### 2.2 机器学习的常见方法

#### 2.2.1 监督学习

监督学习表示机器学习的数据是带标记的, 这些标记可以包括数据类别, 数据属性以及特征点位置等。这些标记作为预期效果, 不断来修正机器的预测结果。具体首先过程是: 通过大量带有标记的数据来训练机器, 机器将预测结果与期望结果进行比对; 之后根据比对结果来修改模型中的参数, 再一次输出预测结果; 再将预测结果与期望结果进行比对……重复多次直至收敛, 最终生成具有一定鲁棒性的模型来达到智能决策的能力。

常见的监督学习有分类、回归。分类 (classification) 是将一些实例数据分到合适的类别中, 它的预测结果是离散的。回归 (regression) 是将数据归到一条“线”上, 即为离散数

据生产拟合曲线, 因此其预测结果是连续的。

#### 2.2.2 无监督学习

无监督学习表示机器学习的数据是没有标记的。机器从无标记的数据中探索并推断出潜在的联系。常见的无监督学习有聚类, 降维。

在聚类 (clustering) 工作中, 由于事先不知道数据类别, 因此只能通过分析数据样本在特征空间中的分布, 例如基于密度或是基于统计学概率模型等等, 从而将不同数据分开, 把相似数据聚为一类。

降维 (dimensionality reduction) 是将数据的维度降低。例如描述一个西瓜, 若只考虑外皮色泽、根蒂、敲声、纹理、密度以及含糖率这 6 个属性, 这 6 个属性代表了西瓜数据的维度为 6。进一步考虑降维的工作, 由于数据本身具有庞大的数量和各种属性特征, 若对全部数据信息进行分析, 将会增加训练的负担和存储空间。因此可以通过主成分分析等其他方法, 考虑主要影响因素, 舍弃次要因素, 来平衡准确度与效率。

#### 2.2.3 强化学习

强化学习是带激励的, 具体来说就是, 如果机器行动正确, 将施与一定的“正激励”; 如果行动错误, 也同样会给出一个惩罚 (也可称为“负激励”)。因此在这种情况下, 机器将会考虑如何在一个环境中行动才能达到激励的最大化, 具有一定的动态规划思想<sup>[2]</sup>。例如在贪吃蛇游戏中, 贪吃蛇需要不断吃到“食物”来加分。为了不断提高分数, 贪吃蛇需要考虑在自身位置上如何转向才能吃到“食物”, 这种学习过程便可理解为是一种强化学习。

强化学习最为火的一个应用便是谷歌 AlphaGo 的升级版——AlphaGo Zero。相较于 AlphaGo, AlphaGo Zero 舍弃了先验知识。不再需要人为设计特征, 直接将棋盘上黑白棋子的摆放情况作为原始数据输入到模型中, 机器使用强化学习来自我博弈, 不断提升自己最终出色完成下棋。AlphaGo Zero 的成功证明了在没有人类的经验和指导下, 深度强化学习依然能够出色完成指定任务。

## 3 机器学习的应用

由于近 20 年科技的迅速发展, 机器学习已经成为了一门活跃并且充满生命力的学科。机器学习具备了一定的解决实际生活问题的能力, 通过把机器学习当作是一种基础与服务技术, 和不同领域进行结合碰撞出火花<sup>[3]</sup>。我们首先通过几个具体的例子来进一步了解机器学习优势。

(1) 标注问题。标注问题就是为数据标记类别, 这些数据可以包括图片, 声音, 文字等等。例如下面这张图片中, 包含的动物有猫和狗两种, 因此可以利用多标注将图片标注为“有猫并且有狗”。现实情况中, 需要标注的数据往往以百万万甚至以亿计数, 人力标注就显得很吃力了。例如在有监督学习中, 大量的数据是需要标注的。因此通过机器学习的自动标注可以节省大量时间成本, 提高工作效率。

(2) 搜索与排序。搜索与排序研究的是如何将一堆对象进行排序。例如在检索信息时, 我们常常关注 (下转第 112 页)

地得到大量的反馈数据,电动汽车共享服务公司并以此为资料来设计专门的调运模型。同时,能够进行实时的数据监控,监控一些共享站点的电动车数量,汽车停车位和空闲电动汽车的数量,以及共享站点空余停车数量的变化和后续订单的情况等。比如对订单数据的监控当中,电动汽车共享服务公司可以实时的获得以下信息:用户(会员)的ID,取车时间地点;还车的时间地点;使用的里程;车辆实时位置等基本数据,同时还可以利用这些基本数据,为基础汽车共享站点之间的车辆进行优化调整。

#### 4.2 优化驾驶体验,节省电力

以车联网技术作为支撑,可以使电动汽车服务公司通过信息端获取到电池、电量等,另外通过ABS系统获取相关信息。借助于信息收集端来获得共享电动汽车的速度信息,包含状态的诊断、车辆状态、加速和减速等相关信息,借助于互联网数据处理技术,让会员可以享受到节能驾驶的驾驶体验。同时也能够通过这些信息使会员更好的驾驶汽车,减少会员在驾驶的过程中由于不合理的驾驶习惯而对电动汽车造成续航里程减少的问题,节省电力,促使电动汽车能够更长时间的节能行进。

#### 4.3 对行车路径进行规划

在电动汽车共享服务当中,用户(会员)在对车辆进行预约出行时,提供地图导航是基本的共享服务之一,包括了

基本的电动汽车共享站点位置的分布查询,实时的路况,充电桩查询,对车辆的导航以及相关的路径规划等服务。同时为了能够及时满足电动汽车共享用户的出行要求,在用户预定汽车时,还能够向用户提供电动汽车的位置信息,并在用户行进的时候,为用户提供基本的导航服务。因为现在电动汽车最大的问题就是汽车的续航里程问题,电动汽车毕竟不能像燃油汽车那样快速的完成燃油补给,因此要对用户提供共享站点的充电桩位置的相关信息,当用户所驾驶的车辆的续航里程低于一定值得时候,给用户发送提示信息,并且推送最近充电桩位置。

### 5 结束语

当前互联网技术已经成为了不少行业中不可或缺的重要组成部分,而车联网则是基于物联网技术形成的一种服务于汽车行业的新兴技术。随着人们对保护环境意识的提升,共享电动汽车以其绿色节能的特点得到不少人的青睐,必将在未来的城市交通发展当中成为主要的发展方向,对此要积极地发展车联网技术以及解决电动汽车的续航里程问题,实现绿色出行和环境保护的城市建设目标。

#### 参考文献

- [1] 张志祥,李静林.电动汽车运营支撑环境研究与设计[J],2013.

(上接第105页)了医疗论文与报道,临床指南,药物使用报告以及上千上万本患者病历等。之后利用AI算法进行分析处理,从而为医生提供医疗辅助,更高效精准地进行医疗诊断。

在零售行业,线下的实体零售店利用AI实现真正的无人零售,降低成本的同时也大大提高了效率。电商巨头亚马逊搭建的智能实体零售店AmazonGo,在极短的时间内为智能零售添了一把火。AmazonGo中一项名为“Just Walk Out”的技术结合了机器学习、计算机视觉和传感器等,通过在店内遍布传感器、摄像头以及信号接收器,可以监测货架上商品的取下与放回,并在虚拟购物车中追踪。而线上,人工智能在推荐系统上的应用将提高在线销售,更精确地实现市场预测,从而降低库存成本。推荐系统针对用户潜在偏好,建构在线商品推荐模型,目前已经在很多电商网站上得到应用。

在传媒行业,通过内容传播机器人以及品牌传播机器人,一键生成用户想要的内容,一分钟内可出一万篇文章,并且进行有效的传播。基于人工智能搭建的智能媒体平台,可以通过研究媒体的投递和发送规则,结合当前热点事件、舆论和公关营销内容,自动生成用户想要阅读的内容。并且可以智能地对接主流媒体平台,自动同步发送,实现了高效传播。在品牌宣传方面,智能平台根据品牌内容、推广预算、推广

效果来预估和匹配媒体产品、媒体渠道,从而赋予了企业最大的传播价值。

### 4 结束语

显然,目前人工智能已经成为了科技界的热门研究方向,各大巨头公司例如谷歌、微软、IBM以及国内的百度、腾讯和阿里巴巴等,都将人工智能作为发展重点,致力于将人工智能应用到更多的领域上。能够结合人工智能的行业多到数不尽,人工智能势必会重新定义人类的工作范围并重塑人类社会,引起一番科技改革大浪潮。

#### 参考文献

- [1] 朱健.利用GPU加速人工智能新型计算模式.计算机与网络,2016.42(2):18-19.  
[2] 王远桂,何欢.人工智能2.0给自动驾驶发展带来的影响.现代电信科技,2017.47(4):20-24.  
[3] 李彦宏.迎接新时代内容分发、连接服务、金融创新、人工智能.市场观察,2017(1):60-65.  
[4] 孔祥溢,王任直.人工智能及在医疗领域的应用.医学信息学杂志,2016.37(11):1-5.

(上接第109页)如何把一堆对象按照检索目的相关性进行排序。在互联网时代,由于各种搜索引擎的流行,我们将目光聚集在如何对网页进行相关性排序。基于机器学习的网页排序,可以根据用户喜好智能并且个性化呈现出搜索结果。

(3) 推荐系统。推荐系统与排序问题紧密相连,并被广泛应用于新闻网站,购物网站以及搜索引擎上。推荐系统的目标是将不同用户各自感兴趣的内容推荐给他们<sup>[4]</sup>。例如Comcast公司(美国一家做宽带网络,有线电视及IP电话服务供应商的公司),利用机器学习技术分析了大量用户的历史记录,对每一个用户输出对应的品味描述,并根据品味描述把顾客分成不同类别。对于不同类别的顾客,Comcast都能实时跟踪并推荐相关的最流行的节目内容。基于机器学习生

成了更好的推荐系统,达到了更高的利用率,促进了更高的顾客满意度。

#### 参考文献

- [1] 吴康宁.基于人工智能下的机器学习历史及展望研究[J].科技商品,2017(6):187.  
[2] 高阳,陈世福,陆鑫.强化学习研究综述[J].自动化学报,2004,30(1):86-100.  
[3] 陈春霞.浅析机器学习的发展与应用[J].信息系统工程,2017(8):99-100.  
[4] 刘忠宝.机器学习方法在个性化推荐系统中的应用[J].情报探索,2016(4):80-82.