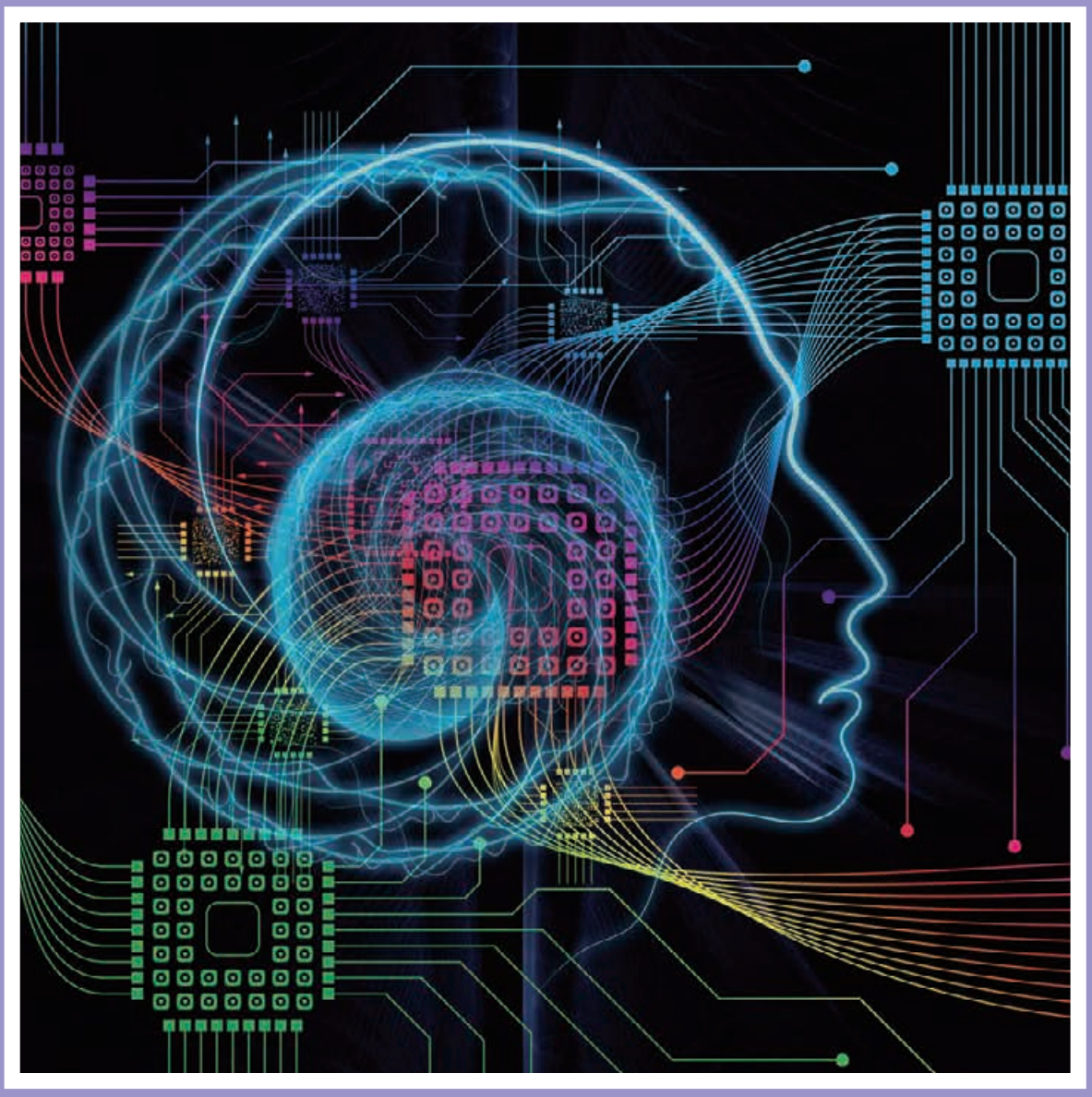
普通高中教科书

信息技术

选择性必修 4

人工智能初步



上海科技教育出版社

普通高中教科书

信息技术

选择性必修 4

人工智能初步

上海科技教育出版社

编写人员名单

主　　编：郑　骏

分册主编：杨小康

主要编写人员（以姓氏笔画为序）：

宋　利　倪冰冰　徐　奕

翟广涛

欢迎广大师生来电来函指出教材的差错和不足，提出宝贵意见。

上海科技教育出版社地址：上海市闵行区号景路 159 弄 A 座 8 楼

邮政编码：201101

联系电话：021-64702058

邮件地址：office@sste.com

写给学生的话

亲爱的同学：

不知你是否留意，近年来生活中涌现了许多与科技相关的新名词：无人驾驶、刷脸支付、城市大脑、阿尔法狗、聊天机器人、智能语音助手、机器学习、深度学习、人工神经网络……其实，这些名词都与你将要学习的人工智能相关，它们有的指人工智能的应用，有的指人工智能的产品，有的指人工智能的算法。

在《人工智能初步》的学习中，我们将为你揭开人工智能的神秘面纱，带你一起走进人工智能的世界。你将探索人工智能的起源和发展历程，思考未来的发展之路；你将了解当前人工智能系统背后的原理，领略其中蕴含的基本思想和方法；你将尝试开发简单的人工智能系统，体验人工智能开发的基本过程与方法。

为了让你在学习《人工智能初步》的过程中获得更大的成功，请浏览本书的栏目介绍。

单元引言、学习目标和单元挑战

从生活经验出发引入本单元将要学习的内容，提出本单元学习要达成的学习目标，预告学习完本单元后要接受的单元挑战。

项目引言和项目学习目标

描述项目产生的背景和意义，介绍项目学习的主要内容，并提出一些具体问题，引导你带着问题探究。

项目学习指引

通过剖析真实的项目实施过程，帮助你了解学科思想方法，理解相关概念，掌握具体技能。

核心概念和小贴士

解释一些重要概念和术语，或提示相关知识和技术，帮助你抓

住重点，扫除认知障碍。

思考与讨论

提出若干问题引导你对技术背后的原理以及人、信息技术与社会的关系等进行思考和讨论。

数字化学习

引导你利用网络、数字化工具和数字资源进行学习。

活　动

提出活动任务，并引导你运用所学知识，使用信息技术工具进行探究、总结和展示。

知识链接

系统整理和归纳本项目的知识要点，方便你学习。

拓展阅读

补充更丰富的阅读材料，开阔你的视野。

单元挑战

布置面向真实情境的项目任务，希望你综合运用本单元所学的知识与技能去解决问题。

单元小结

用思维导图可视化呈现本单元的知识脉络，提供基于学科核心素养的评价表，为你的学习表现进行自我评价。

在学习过程中，希望你勤实践体验、多思考讨论，借助各种数字化工具、资源进行学习与创新，不仅要理解和掌握具体的信息技术知识与技能，还要把握用信息技术解决问题的思想方法，并思考将信息技术应用于社会时所引发的各种挑战，以开放、包容的心态与信息技术、信息社会一起进步。

编　　者

目 录

第一单元 走进人工智能的世界 1
项目一 初识人工智能——了解人工智能的发展历史与现状 2
1. 通过实例感受人工智能 3
2. 回顾人工智能的发展历程 5
3. 了解当前人工智能技术应用状况 9
知识链接 10
项目二 探秘智能车——认识人工智能系统 12
1. 了解人工智能系统的基本技术要素 13
2. 分析智能车的机构组成 16
3. 明确人工智能的特征 20
知识链接 21
单元挑战 探究服务机器人 23
单元小结 24

第二单元 理解人工智能技术的思想与方法 25
项目三 让智能车能够“刷脸”开车门——探究图像识别与理解 26
1. 了解图像识别与人脸识别技术 27
2. 获取、表示人脸特征 29
3. 解析人脸识别原理 33
4. 评价人脸识别性能 37
知识链接 38
项目四 让智能车与用户对话——探究语音交互技术 42
1. 认识语音交互 43
2. 让机器理解语音 45
3. 用算法实现语音识别 47
4. 应用语音交互技术 48
知识链接 50
项目五 让智能车自动规划路径——探究智能决策与搜索算法 52
1. 用人工智能实现路径规划 53
2. 预测交通流量 54
3. 根据路况进行智能决策 56
4. 搜索最佳路线 59

知识链接 63
项目六　让智能车识别道路障碍物——认识人工神经网络与深度学习 66
1. 初识人工神经网络 67
2. 了解深度学习及其基本操作 70
3. 探索深度学习的最新发展 73
知识链接 77
项目七　在车展中实现“车以类聚”——探究无监督学习与聚类算法 82
1. 认识无监督学习与聚类算法 83
2. 剖析 k- 均值聚类算法 84
知识链接 90
单元挑战　用 SVM 算法及深度学习给图像分类 92
单元小结 93

第三单元　开发简单人工智能系统 95
项目八　搭建可“刷脸”启动的循迹智能车——设计简单的人工智能系统 96
1. 进行总体设计，确定基本开发方案 97
2. 设计人脸识别启动系统 98
3. 设计智能车循迹系统 100
4. 测试智能车 105
知识链接 106
单元挑战　设计智能车避障系统 109
单元小结 110

第四单元　推动人工智能健康发展 111
项目九　认识人工智能的巨大价值和潜在威胁——辩证看待人工智能 112
1. 了解人工智能技术的应用现状 113
2. 直面人工智能的安全、伦理问题 115
3. 展望人工智能的未来 119
知识链接 121
单元挑战　设计无人驾驶时代的交通准则 123
单元小结 124

附录　部分名词术语中英文对照 125

第一单元

走进人工智能的世界

2016 年，AlphaGo(一款下围棋的人工智能程序) 的横空出世掀起了人工智能的新一轮热潮，人工智能再一次成为社会各界关注的焦点。这距离 1956 年人工智能概念首次提出已经有 60 年了。人工智能的发展由于受到智能算法、任务相关数据以及机器计算能力等因素影响，历经了多次起伏。直到 2006 年，以深度学习为代表的智能算法在计算机视觉和自然语言处理等领域取得了重大突破，同时，大数据、云计算等技术为人工智能的发展提供了丰富的数据资源与计算资源，人工智能才开始全面爆发，在众多应用领域都取得了极大的成功。

在本单元中，我们将走进人工智能的世界，了解人工智能的发展历史、重要事件及其在现实世界中的应用。同时，我们将通过剖析典型的人工智能系统，认识人工智能的要素以及智能系统的组成。

单元挑战

探究服务机器人

学习目标

- 能描述人工智能的概念与基本特征。
- 知道人工智能的历史、典型应用与发展趋势。

项目一

初识人工智能

——了解人工智能的发展历史与现状

棋类运动一直被看作人类的高智商游戏，围棋更被视为人类最后的智慧堡垒。然而在过去短短 20 年内，人类在各种棋类运动中相继被人工智能击败。1997 年，“深蓝”超级计算机战胜了当时排名世界第一的国际象棋大师卡斯帕罗夫。2006 年，5 位中国象棋特级大师与超级计算机“浪潮天梭”展开对决，最终败给了计算机。2016 年，AlphaGo 以 4 比 1 的比分战胜了韩国围棋棋手李世石；仅仅过了一年，升级版的 AlphaGo 又以 3 比 0 击败了当时排名世界第一的中国围棋棋手柯洁（图 1-1）。

AlphaGo 的成就再度点燃了人们对人工智能的热情。放眼望去，在现实世界中，智能视频监控、自动对话机器人、自动装配机器人、无人机等各种人工智能系统已经广泛应用于生产和管理的多个领域。这也在告诉我们，人工智能的春天到来了！

图 1-1　柯洁与 AlphaGo 对弈

项目学习目标

在本项目中，我们将了解什么是人工智能，回顾人工智能的发展历史，探索当前人工智能技术的应用状况。

完成本项目学习，须能回答以下问题：

1. 什么是人工智能？
2. 人工智能学科是怎样诞生的？
3. 人工智能的发展经历了哪些阶段？
4. 人工智能发展各阶段的特征是什么？

项目学习指引

学会制造与使用工具，是人类发展历程上的重要突破。让机器具备“人类的智能”，是人类自古以来追求的梦想。从中国古代的“偃师人偶”的传说，到希腊神话中的机械人和人造人，无不体现人类对于制造出智能机器的渴望。现代意义上的人工智能，始于哲学家与数学家用机械符号处理的观点解释人类思考过程的尝试。1956 年的达特茅斯会议，标志着现代人工智能作为一门学科的诞生。

1. 通过实例感受人工智能

人工智能是当前非常热门的研究领域。经过几十年的发展，目前人工智能的研究涵盖了从感知、学习、决策等通用方法到自动驾驶、人脸识别、医疗辅助诊断等专门领域。那么，什么是人工智能呢?

（1）从自动驾驶系统看环境感知

汽车司机要想做到安全熟练的驾驶，不仅仅要准确识别道路上的各种标识与符号，更重要的是在行驶过程中，对周围的行人和车辆等动态物体的行为进行预判，据此进行决策。这种预判能力在人流密集的道路上尤为重要。人类驾驶员在行车过程中会对行人的移动轨迹进行预判，以保持一个安全距离。以此类比，一个自动驾驶系统若想要代替人类驾驶员，不仅需要识别出静态的交通标识与物体，对周边行驶环境进行准确的感知，还要对时刻变化的路况进行预判。

环境感知，是自动驾驶汽车认路的关键组成部分。自动驾驶系统中用于环境感知的部件有许多种类，如视频摄像机、激光测距仪、车载雷达、速度传感器等多种车载传感器。自动驾驶系统依靠这些传感器收集周围的路面环境信息，供决策系统进行分析，进而作出相应的路径规划，如图 1–2 所示。

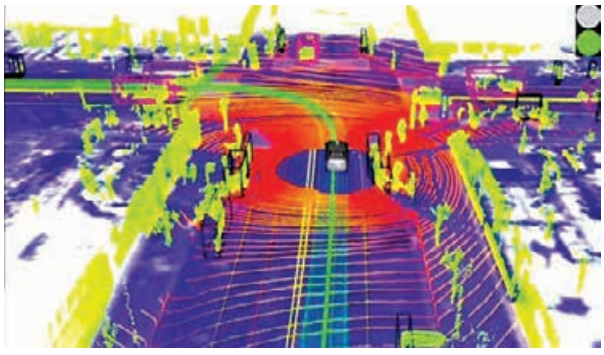


图 1–2　某自动驾驶汽车利用车载传感器“看”到的环境

核心概念

人工智能（Artificial Intelligence，简称 AI）是一门前沿交叉学科，对其定义一直存有不同的观点。一般认为，它是利用计算机或者计算机控制的机器，模拟、延伸和扩展人的智能，感知环境、获取知识并使用知识获得最佳结果的理论、方法、技术及应用系统。

思考与讨论

各种车载传感器是自动驾驶汽车感知外部环境的关键。这些传感器分别采集哪些类型的信息？

环境感知不仅是自动驾驶系统的关键组成部分，还是无人机、对话机器人等人工智能系统中的重要部分。人工智能系统要体现出“智能”，必须先感知周围的情况，这样才能进一步作出相应的决策。比如智能翻译系统只有先采集并感知语音信息，才能进行后续的翻译处理。

（2）从答题机器人看智能决策

在各类电视节目中，一些答题类的综艺节目往往会吸引大量的参赛者与观众。在现实中，人工智能系统也曾作为参赛者参与过这类综艺节目。2011 年，一款答题机器人——沃森系统（Watson）参与了答题类综艺节目《危险边缘》，并击败人类冠军选手，如图 1-3 所示。节目中，沃森的抢答速度始终快于人类选手，对于绝大多数问题均可以给出正确答案，展现了它的决策速度与决策能力。

图 1-3　沃森系统参与答题类综艺节目

环境感知是一个智能系统拥有“视觉”“听觉”“触觉”等知觉能力的基础，而智能决策部分则可看作是一个智能系统能够进行深度思考的核心。沃森系统拥有这样强大的智能决策能力：对于用双关语提出的甚至对很多观众而言都很费解的问题，它依然能作出分析和推理，并在巨大的自然语言数据库中寻找线索，然后将这些线索合成答案，最后用自然语音进行回答。

思考与讨论

你认为智能答题系统需要哪些关键部分来完成“听到题目—作出判断—回答问题”的过程？

（3）认识人工智能和人工智能系统

人工智能一般定义为研究和开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的一门技术科学。

在现实生活中，人们总是主动或被动地感受周围的环境并作出相应的反应。例如看到一只猫，大多数人会下意识地觉得可爱，产生想去抚摸的愿望；又如听到有人叫唤自己的

名字，便会情不自禁地转头去寻找声音来源。与此对应，可把人工智能系统看成是一个能够进行环境感知，并根据采集到的信息作出决策以达到特定目标的机器系统。自动驾驶系统和沃森系统就是这样的人工智能系统，它们能够感知外界信息，并根据信息作出决策，从而实现自动驾驶和自动答题。

2. 回顾人工智能的发展历程

（1）人工智能的诞生（1956 年）

人工智能在正式诞生之前，经历了漫长的孕育期，这甚至可以追溯到 17 世纪莱布尼茨发明的二进制表示规则。到 20 世纪 40 年代，一些研究人员已经开始探索如何让机器具有智能。

1950 年，阿兰·图灵（Alan Turing）提出著名的图灵测试（The Turing Test）：如果一台机器能够与人类展开交互而不被辨别出其机器身份，那么就称这台机器具有智能。图灵测试被认为是检测一台机器是否具有智能的重要标准。

参见 P10 知识链接“图灵测试”

1951 年，马文·明斯基（Marvin Minsky）等人搭建了第一个神经网络模拟器——SNARC（随机神经网络模拟加固计算器），它使用 3000 个真空管来模拟 40 个神经元（neuron）的活动。虽然当时还没有“人工智能”这个概念，但这项开创性工作为后来的人工智能发展奠定了深远的基础。

1956 年夏天，约翰·麦卡锡（John McCarthy）、马文·明斯基、克劳德·香农（Claude Shannon）与内森尼尔·罗切斯特（Nathaniel Rochester）等多位科学家在美国达特茅斯学院里组织了一个为期两个月的研讨会，研究让机器来模拟智能的可能性。在会上，麦卡锡首次提出“人工智能”的概念，这被认为是人工智能正式诞生的标志。参会的赫伯特·西蒙（Herbert Simon）和艾伦·纽厄尔（Allen Newell）展示了他们编写的推理程序——“逻辑理论家”（Logic Theorist），该程序因可以证明《数学原理》中的多个定理而受到了高度关注。正是这些与会人员的深入讨论与集思广益，使得人工智能成为一门独立的学科。因此，1956 年被视为人工智能元年。随后，人工智能的发展经历了一段曲折的过程，如图 1–4 所示。

（2）第一个黄金期（1957—1974 年）

人工智能概念的诞生让人们看到了使机器具有智能的可能性。研究人员开始以极大的热情开展人工智能相关领域的

小贴士

达特茅斯学院研讨会的提案声明：

我们提议 1956 年夏天在新罕布什尔州汉诺威镇的达特茅斯学院开展一次由 10 个人组成的为期两个月的人工智能研究。这项研究基于这样的推测：原则上可精确地描述学习对象的每个方面或智能体的任何特征，从而能够建造一台机器来模拟它。该研究将尝试发现如何使机器使用语言，形成抽象与概念，求解多种现在注定由人来求解的问题，提升自我（机器）。我们认为：如果仔细选择一组科学家对这些问题一起研究一个夏天，那么对其中的一个或多个问题就能取得意义重大的进展。

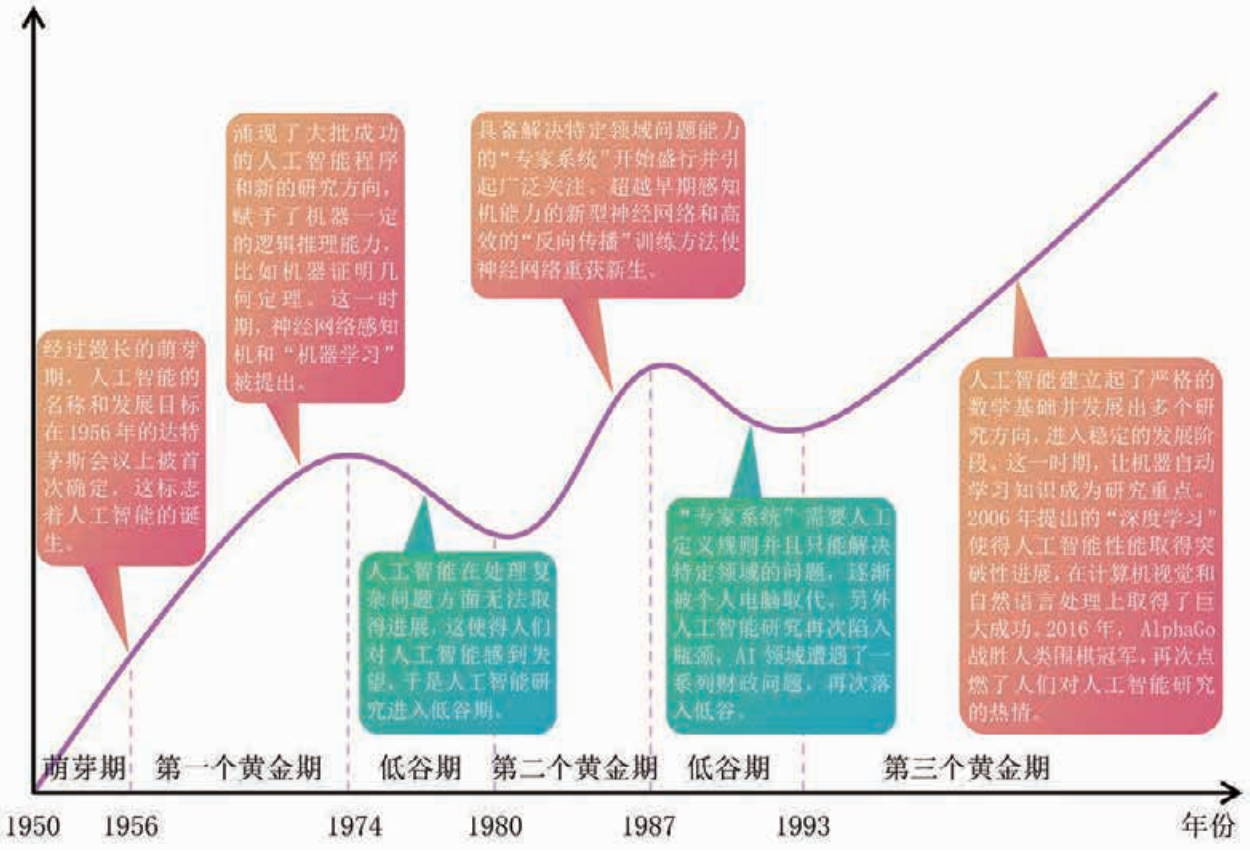


图 1-4　人工智能发展历程

核心概念

机器学习是一种让机器从数据或者行动中学习以获得预测或判断能力的方法。

小贴士

20 世纪 70 年代后期，我国数学家吴文俊提出用计算机证明几何定理的“吴方法”，开创了现代数学史上第一个由中国人原创的研究领域——数学机械化，实现了将繁琐的数学运算证明交由计算机来完成的目标，使得机器在相关问题上具有了与人类相似甚至比普通人更强的推理能力。

研究。很快，人工智能就迎来了第一个发展高峰期。这一时期的人工智能研究认为只要赋予机器一定的逻辑推理能力就可以实现机器智能。1957 年，纽厄尔和西蒙在“逻辑理论家”的基础上发明了“通用问题求解器”（general problem solver），它通过模仿人类求解问题的逻辑来解决问题。1959 年，赫伯特·格伦特（Herbert Gelernter）建造了一个几何定理证明器，能够自动证明一些数学专业学生都感到棘手的定理。

除了让机器具有逻辑推理能力的研究外，许多研究人员尝试通过其他方法实现机器智能。比如，1957 年，康奈尔大学的弗兰克·罗森布拉特（Frank Rosenblatt）完成第一个神经网络模型——感知机，实现了识别（或分类）输入信号中的简单图案。1959 年，阿瑟·萨缪尔（Arthur Samuel）提出了“机器学习”（Machine Learning，ML），并将机器学习定义为让计算机不需显式编程也可以自动学习的领域。1966 年，MIT 的约瑟夫·魏泽鲍姆（Joseph Weizenbaum）展示了早期最知名的人机对话程序——Eliza，它可以模拟心理医生与人类展开简单的对话。

虽然这段时期涌现了许多的相关成果，一些简单的推理

任务也可以通过机器智能程序自动完成，但在稍显复杂的问题上，人工智能一直无法取得进展。主要原因在于当时的计算机计算性能不足，受限的内存容量和处理速度导致计算机程序无法解决复杂问题。由于数据的严重缺失，计算机程序无法从数据中学习到足够的知识。因此，研究进展逐渐减慢，当初许多科学家的预期迟迟无法实现，人们开始对人工智能感到失望。许多机构停止了对人工智能研究的资助，人工智能进入了一段艰难的低谷期。

（3）第二个黄金期（1980—1987 年）

随着研究的推进，人们逐渐意识到，仅仅使机器具有推理能力是实现不了人工智能的。要想真正使机器具有智能，一个重要前提是要让机器具有知识。正是在这种思想的指导下，专家系统诞生了。它是人工智能三个主要研究流派中的符号主义流派的代表性系统。

参见 P11 知识链接“人工智能主要流派”

第一个实用的专家系统是 1969 年发布的 DENDRAL，用以帮助化学家推断物质的分子结构。进入 20 世纪 80 年代后，专家系统快速发展，使得人们对人工智能的热情再度高涨，人工智能迎来了第二个黄金期。当时著名的专家系统之一是由卡内基梅隆大学在 1980 年为 DEC 公司设计的 XCON 系统，它可以根据用户的需求自动组合配件，为用户组装计算机。该专家系统为 DEC 公司显著地节省了开支。

与此同时，在连接主义研究流派中，神经网络方面的研究进展也让人们重新看到了人工智能的潜力。1957 年提出的感知机，结构过于简单，能力非常有限。1982 年，加州理工学院的约翰·霍普菲尔德（John Hopfield）提出一种新的神经网络，可以用来解决多种模式识别问题，神经网络研究界因此振奋。由于不知道如何有效调整中间层的网络结构参数，早期的神经网络往往比较简单。1986 年，戴维·鲁梅尔哈特（David Rumelhart）、杰弗里·辛顿（Geoffrey Hinton）和罗纳德·威廉姆斯（Ronald Williams）利用反向传播（Back Propagation，BP）算法，较好地解决了大规模神经网络的训练问题。

然而，到 20 世纪 80 年代初，个人计算机出现了，因其价格远远低于专家系统等人工智能系统，且通用性远超专家系统，人们对人工智能系统的热情开始下降。专家系统需要人工定义规则，这项工作不但费时费力，而且在语音识别、图像识别等自然输入的应用场合中难以实施。专家系统很多功能很容易被个人计算机的通用软件所替代。20 世纪 80 年

小贴士

专家系统（Expert System，ES）是一类具有某个领域内专家水平的知识与经验的智能计算机程序系统。专家系统根据领域内一个或多个专家提供的知识和经验，模拟人类专家的决策过程进行推理判断。简而言之，专家系统是一种模拟人类专家解决其领域内相关问题的计算机程序系统。

小贴士

神经网络（Neural Network，NN）指人工神经网络，是模仿活体生物体系统神经元网络的模型，用来模拟人类大脑神经系统的结构和功能。

这是人工智能中一个重要的研究领域。

代末期，因美国国防高级研究计划局（DARPA）的人工智能计算机没能达到研究目标，美国政府开始缩减对人工智能方向的投入，将资助转向了其他更容易出成果的项目。人工智能进入第二个低谷期，大量的人工智能公司倒闭。

（4）第三个黄金期（1993 年至今）

1993 年开始，人工智能研究逐渐走出“寒冬”。这一时期，研究人员逐渐建立起人工智能的严格数学基础，人工智能转变成一门严格的科学分支。人工智能的研究领域不断扩大，形成了专家系统、机器学习、计算机视觉、自然语言理解等方向。人工智能研究逐渐到达一个稳定的阶段，研究重点由教给机器某领域内的特定知识变为让机器自动学习知识。

作为一种让机器自动学习知识的重要方式，机器学习从人工智能诞生起就一直是研究的重点。机器学习的基本思想是让机器从数据或者行动中学习，获得进行预测或判断的能力。从数据中学习，指利用算法从大量的训练数据中学习知识，并通过学习不断优化程序的性能，然后用经训练而优化过的程序对真实世界中的待测试数据作出判决。典型的算法包括决策树、聚类、贝叶斯分类、神经网络等。从行动中学习，指智能体在跟环境的交互过程中，根据回报情况来学习一套指导行动的策略。这样的机器学习方式称为强化学习（reinforcement learning）。小孩学走路、学下棋的过程是典型的强化学习，AlphaGo 系统超越人类的围棋博弈能力也主要是通过强化学习训练出来的。

2006 年，杰弗里 · 辛顿等人提出深度学习方法，使得人工智能研究取得突破性进展。神经网络的隐藏层越多，学习能力越强，但计算的复杂度也急剧增加。先前的神经网络一般只有一到两层隐藏层，辛顿等人提出逐层预训练以及降维的方法，减少了对人类先验知识的依赖，使得神经网络学习知识的能力显著提升，引发了研究深度学习的浪潮。深度学习在人工智能的诸多领域取得巨大成功。以车牌检测为例，运用深度学习可以实现自动学习矩形的车牌形状、突出的数字 / 字母、不同的底色、车牌规定的尺寸等特征，性能显著优于基于人工特征的方法。

2012 年，得益于人工智能 ABC 三要素［先进的深度学习算法（Algorithm）、海量的数据（Big data）以及强大的计算能力（Computing power）］的发展，人工智能开始突飞猛进。2013 年，深度学习在图像识别和语音识别领域取得突破，

小贴士

深度学习（Deep Learning，DL）是一系列算法的统称。深度学习算法通过组合多层的神经网络，来模拟人脑在处理数据时由底层到高层的抽象过程。

与传统神经网络等相关方法相比，深度学习的网络层数更多、网络规模更大、学习能力更强。深度学习是机器学习的重要分支。

标志着人工智能实现了感知智能。

2016 年，AlphaGo 横空出世，它借助深度学习及先进搜索算法的强大威力，横扫围棋界，攻克棋类运动中人类最后的智慧堡垒。人工智能领域中又一次掀起了新的浪潮。

3. 了解当前人工智能技术应用状况

自深度学习出现后，人工智能开始渗透到生产生活的方方面面，赋予各行各业新的能量。它降低了生产成本，加快了产品与技术的迭代速度，提高了经济效益和社会效益。在我国，人工智能在众多行业和领域中得到广泛应用，图 1–5 呈现的是 2018 年 13 个行业中 61 个领域的人工智能技术应用商业化程度及技术应用深度的情况。

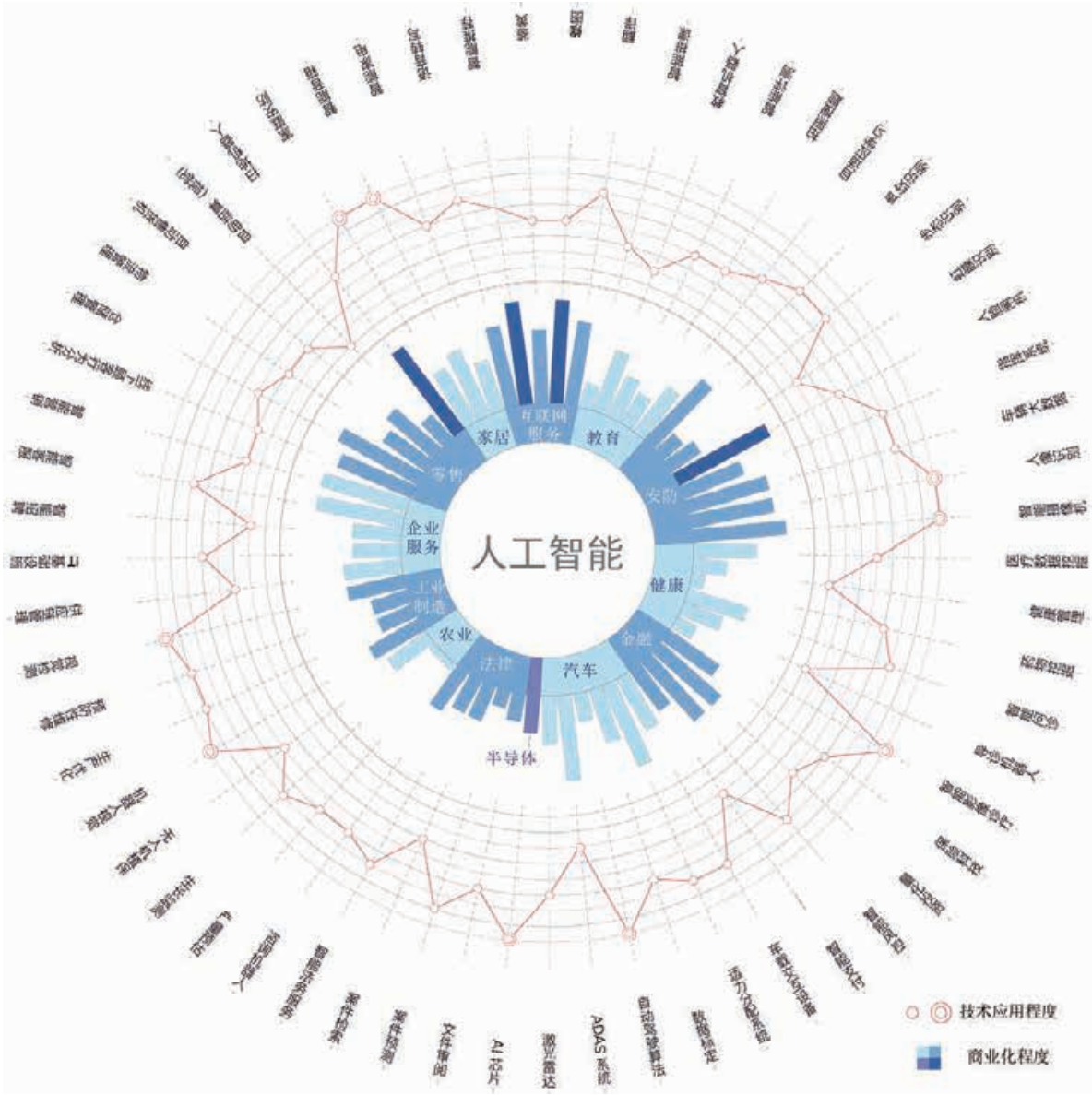


图 1–5　人工智能在各领域的应用情况（摘自亿欧智库《2018 中国人工智能商业落地研究报告》）

人工智能已成为国家的重要发展战略。2016—2017 年，我国政府陆续发布了《“互联网 +”人工智能三年行动实施方案》《新一代人工智能发展规划》《促进新一代人工智能产业发展三年行动计划（2018—2020 年）》等政策文件，促进人工智能技术的发展，培育人工智能新兴产业，鼓励人工智能化创新创业，带动我国产业升级和经济转型。我国政府还确定了百度“自动驾驶”、阿里云“城市大脑”、腾讯“医疗影像”、科大讯飞“智能语音”四个首批国家新一代人工智能开放创新平台，期望它们在汇聚创新资源、促进众创共享方面发挥更大的作用。

活　动

1.1 在人工智能发展历程图中选择其中的一段时期，查找相关资料，串讲人工智能在该时期内的发展历程。要求：阐明这段时期的人工智能发展的特征；介绍该段时期内发生的代表性事件以及这些事件背后的原理技术。

知识链接

图灵测试

图灵测试是由阿兰·图灵在 1950 年发表的著名论文《计算机器与智能》中提出的用于判断“机器是否具有智能”的方法。

简单来说，图灵测试是这样的过程：被测试者［包括一台被测试的机器（图 1-6 中的 A）和一个人（图 1-6 中的 B）］与测试人员（图 1-6 中的 C）隔离，然后测试人员通过一些装置向被测试者随意提问。如果测试人员在问完全部问题后不能判断被测试者中哪个是机器哪个是人，那么这台机器（A）就通过了测试，被认为具有智能。

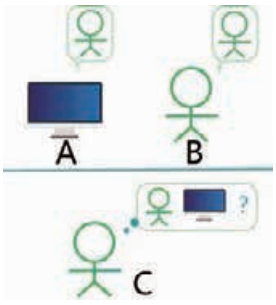


图 1-6　图灵测试图解

人工智能主要流派

在人工智能的研究发展期间，不同学科或学科背景的学者对人工智能提出了各自的理解及不同的观点，由此产生了不同的学术流派。其中，对人工智能研究影响较大的主要有符号主义、连接主义和行为主义三大流派。

（1）符号主义，又称逻辑主义、心理学派或计算机学派，其核心思想是使用符号、规则和逻辑来表达知识并进行推理，代表成果是专家系统。

（2）连接主义，又称仿生学派或生理学派，其核心思想是利用神经网络之间的连接机制模拟智能。深度神经网络是其典型代表。

（3）行为主义，又称进化主义或控制论学派，它是一种基于“感知—行动”的行为智能模拟方法。行为主义认为人工智能源于控制论，推崇控制及感知系统，主要成就是智能控制和智能机器人系统。

拓展阅读

吴文俊与数学机械化

图 1-7　吴文俊

吴文俊（1919 年 5 月 12 日—2017 年 5 月 7 日，图 1-7），浙江嘉兴人，出生于上海。1940 年毕业于上海交通大学，1949 年获得法国斯特拉斯堡大学博士学位。我国著名数学家、人工智能专家。

吴文俊的研究工作涉及数学的诸多领域，其主要成就表现在拓扑学和数学机械化两个领域。他为拓扑学做了奠基性的工作，他的示性类和示嵌类研究被国际数学界称为“吴公式”“吴示性类”“吴示嵌类”，至今仍被国际同行广泛引用。

20 世纪 70 年代后期，在计算机技术大发展的背景下，他继承和发展了中国古代数学的传统（即算法化思想），开始研究几何定理的机器证明，彻底改变了该领域的面貌。他的研究是国际自动推理界先驱性的工作，被称为“吴特征列方法”，产生了巨大影响。

他在拓扑学、自动推理、机器证明、代数几何、中国数学史、对策论等研究领域均有杰出的贡献，在国内外享有盛誉。他的“吴方法”在国际机器证明领域产生了巨大的影响，有广泛而重要的应用价值。当前国际流行的主要符号计算软件都实现了吴文俊的算法。

——摘自科学技术文献出版社《信念　创新　奉献——国家最高科学技术奖获奖者风采》

项目二

探秘智能车

——认识人工智能系统

由于智能算法的突破、数据量的增加、计算能力的提升，服务机器人、智能医疗辅助系统、智能视频监控系统、智能新闻与广告编辑、智能客服、扫地机器人、智能车、无人机等众多人工智能系统已经走进人类的生产和生活。

近年来，智能车（图 1-8）已经成为世界车辆工程领域研究的热点和汽车工业增长的新动力，很多国家都将其纳入到各自重点发展的智能交通系统中。所谓“智能车”，就是在普通车辆的基础上增加了先进的传感器（如车载雷达、摄像头）、车载计算机等装置，使车辆具备智能的车内外态势感知能力，能够自动分析车内外状态，并使车辆按照人的意愿自动行驶到目的地，最终实现替代人来操作的目的。

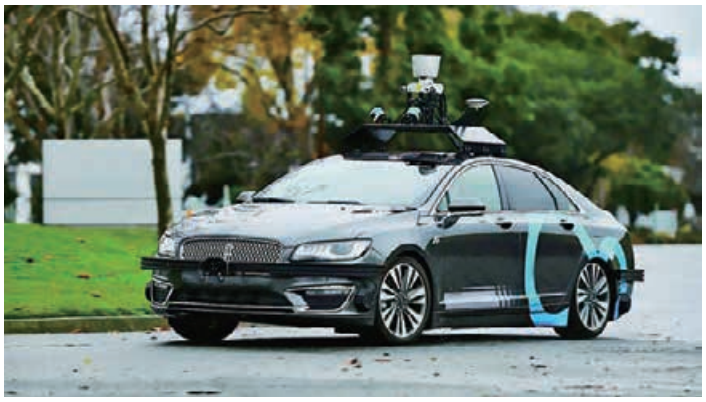


图 1-8　某品牌智能车

项目学习目标

在本项目中，我们将剖析智能车这个典型的人工智能系统，探究人工智能系统的基本技术要素以及人工智能系统的主要组成机构，归纳人工智能的基本特征。

完成本项目学习，须能回答以下问题：

1. 人工智能系统的基本技术要素有哪些？
2. 人工智能系统的主要组成机构有哪些？
3. 人工智能系统的主要组成机构之间有怎样的联系？
4. 人工智能的基本特征有哪些？

项目学习指引

每一个人工智能系统都由基本的技术要素和机构组成，智能车作为一个人工智能系统同样如此。它离不开算法、数据与计算能力等基本技术要素的支撑，也不能缺少感知、决策与执行等组成机构的支持与配合。只有这些基本技术要素与组成机构相辅相成，紧密结合，智能车才能稳定地工作，展现它的智慧和能力。

活　动

2.1 以小组为单位，收集我国大学生智能车竞赛的相关资料，了解大赛的规则，观看历届大赛视频，对智能车设计形成初步印象。

1. 了解人工智能系统的基本技术要素

人工智能目前取得的巨大成就和最新突破离不开以深度学习为代表的先进智能算法、大数据时代带来的海量数据以及以并行计算、云计算为代表的强大计算能力的支持。其中，算法、数据以及计算能力被认为是一个典型的人工智能系统的基本技术要素。

（1）人工智能系统中的典型智能算法

智能算法已经渗透到我们日常生活的方方面面。监控系统中的人脸识别，智能手机上的自动语音识别，网上购物时的商品推荐，音乐软件中的歌单推荐，其背后都是一系列智能算法。

对人工智能而言，智能算法通常包含两个部分，第一是指从海量数据中学习相关知识的方法，第二是用学习到的知识解决实际应用问题的方法。

在人工智能概念提出后，涌现了大批的智能算法，比如决策算法、聚类算法、分类算法、搜索算法、人工神经网络以及深度学习算法等。在众多的智能算法中，掀起最近一次人工智能浪潮并且让人工智能产生广泛落地应用的算法是深度学习算法。

常用的深度学习算法包括卷积神经网络（Convolutional

活　动

2.2 针对不同的人工智能技术（可自行扩充），分组搜索其所涉及的智能算法，并且根据理解列举该技术可能的应用领域，记录于表 1-1 中。

表 1-1　智能算法的应用领域

技术	算法	应用领域
文本分类	例：朴素贝叶斯，支持向量机，决策树	例：新闻分类，垃圾邮件分类，搜索引擎
图像识别		
语音处理		
行为识别		

Neural Networks，CNN）、循环神经网络（Recurrent Neural Networks，RNN）等。深度学习算法在图像识别、自然语言处理等领域取得了极大的成功，比如，采用深度学习算法的人脸识别系统的准确率已经可以超过人类自身的识别准确率。

（2）数据在人工智能系统中的作用

数据是人工智能得以成长的“养分”。人工智能提出后，前期的发展一直比较缓慢，其中一个原因是缺乏足够的数据供算法进行学习，导致算法对实际问题的适应能力一直不强，难以满足人们对人工智能的期待。例如 20 世纪 80 年代提出的神经网络，早期由于训练数据的缺乏，研究人员只能训练小规模的神经网络，无法付诸实用。进入 21 世纪后，有了大量的数据可供大规模的多层神经网络训练时使用，从而使神经网络巨大的能力得以彰显。

直至进入 21 世纪，由于互联网、物联网、数码相机、智能手机等技术和设备的普及，人类开始进入大数据时代，全球数据出现爆炸式增长。研究人员终于可以获得大量的数据来训练复杂的智能算法，使其从海量的数据中学习丰富的知识。比如，在 AlphaGo 系统的学习过程中，核心训练数据是来自互联网的 3000 万例棋谱。

（3）人工智能需要的计算能力

限制人工智能早期发展的另一个原因是当时计算机的计算能力不足以支撑算法获得足够的智能去解决实际问题。随着芯片技术的进步，计算机 CPU 计算能力显著提升，特别是 GPU 在深度学习上的大规模使用，使得人工智能突飞猛进。深度学习中的神经网络在训练时涉及大量可以并行处理的矩阵运算，而 GPU 专门设计的架构拥有成千上万的内核，非常适合多核并行的计算模式，显著加速了深度学习的计算效率。CPU、GPU、FPGA、ASIC 等不同类型的芯片的综合集成，进一步提升了计算能力，具有代表性的是我国自行研制的“神威 · 太湖之光”（图 1-9）和“天河二号”超级计算机系统。

由此看到，计算能力的大幅提升使得人工智能的先进算法能够得以实现，相关大数据能够得以处理。

思考与讨论

“神威 · 太湖之光”超级计算机系统的计算能力怎样？它可以应用于哪些领域？

（4）以智能车为例剖析三个基本技术要素

作为一个典型的人工智能系统，智能车的实现离不开人工智能的算法、巨量的数据和强大的计算能力。

首先看算法。智能车中智能功能的实现需要人工智能算法的支持。比如，要实现“刷脸”上车的功能，必须使用人脸识别算法；要让智能车能够根据采集到的数据理解当前的路况，必须使用图像识别、物体识别等计算机视觉算法；要为当前车辆规划出一条最佳行驶路径，必须使用决策算法。在未来的智慧城市中，人工智能还能统筹分析各个路口的交通状况，为车辆动态地规划最佳路线，在保证顺畅行驶的同时，最大化地发挥道路的运载能力。

其次看数据。智能车在行驶时需要收集巨量的数据，一般每秒钟收集到的数据都是以 GB 为单位。据统计，每台在路上行驶的无人驾驶的智能车每天要处理的数据量为 10 多个 TB，而在上路测试之前更是需要在各种外部环境以及路况下采集数据以进行训练，这种训练数据可达到成百上千个 PB。

小贴士

CPU：Central Processing Unit，中央处理器

GPU：Graphics Processing Unit，图形处理器

FPGA：Field Programmable Gate Array，现场可编程门阵列

ASIC：Application Specific Integrated Circuit，专用集成电路

参见 P21 知识链接“超级计算机”

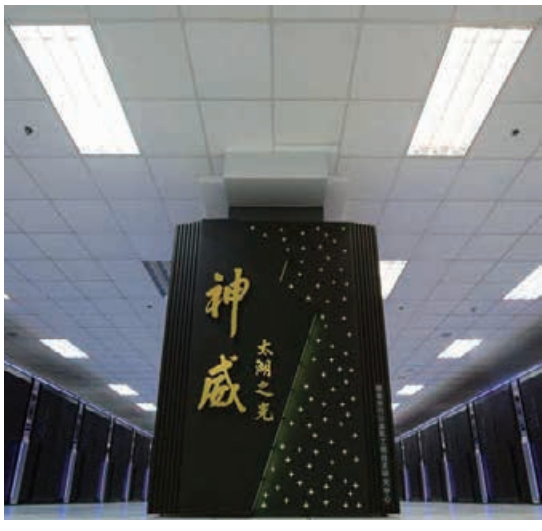


图 1-9　“神威 · 太湖之光”超级计算机系统

参见 P21 知识链接“与智能车相关的人工智能算法”

小贴士

1PB=1024TB；

1TB=1024GB；

1GB=1024MB；

1PB 为 2 万到 3 万部蓝光电影的容量。

最后看计算能力。面对收集到的巨量数据，智能车必须配备足够的计算能力。普通汽车的微控制单元（Micro Control Unit，MCU）已经不能满足智能车数据实时处理的要求。因此，智能车大多采用 CPU、GPU 与 MCU 配合使用的综合控制单元，每种处理器对应处理适合的计算过程。

由此看到，只有算法、数据和计算能力这三大基本技术要素联合发展，才能实现人工智能的突破，才能涌现更多的人工智能系统。

活　动

2.3 上网搜索，了解中国在智能车领域的最新发展。举例分析中国智能车的发展与算法、数据和计算能力提升的关系。

2. 分析智能车的机构组成

智能车的基本工作原理如图 1–10 所示。通过视频摄像头、激光测距仪、车载雷达、速度传感器等传感设备，车辆可以获取行驶状态以及周边路况，分析自身所处的位置、与周围汽车的距离和相对速度等，进而实时规划行驶路径，让汽车能针对不同情况选择不同的应对措施，同时向电机、转向轴发送控制信号，以准确执行规划的行驶策略。在此过程中，摄像头等传感器会不断地将捕获到的画面和车辆运行状态传递给智能汽车的“决策中枢”，从而实时调整智能车的控制策略，以得到最安全、快捷的行车路径。

小贴士

感知机构、决策机构和执行机构是支持人工智能系统稳定工作的三大主要机构组成。

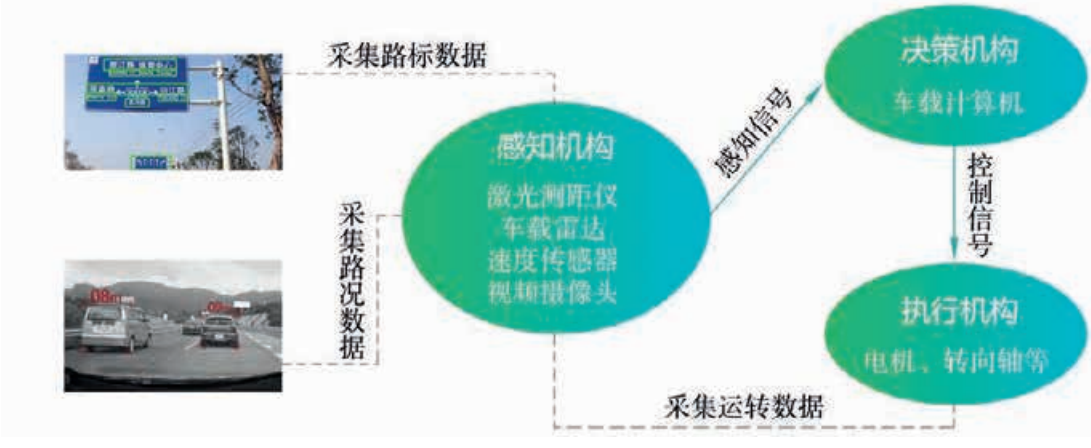


图 1–10　智能车的工作原理

上述过程也体现了智能车的主要组成部分：感知机构、决策机构和执行机构。

（1）智能车的感知机构

人类时时刻刻都在通过耳、鼻、眼等器官来感知、理解外界环境，以进行学习与交流。通过眼睛，观察到自然界的色彩与光线；通过鼻子，感受到自然界的气味；通过耳朵，聆听周围的声音。这些形式多样的信息在大脑中共同交织成完整的环境感知景象，成为认知与决策的基础。

类似于人类，智能车的感知机构（图 1–11）也具有相似的感知功能，但它不是通过眼睛、鼻子、耳朵这类器官，而是通过各种各样的传感设备来实现。例如，外视摄像头可用于侦测交通信号灯，以及行人、自行车等车辆行驶路线上遭遇的移动障碍；内视摄像头可用于感知驾驶员的状态，如是否打瞌睡；车载麦克风可用以识别驾驶员的控制指令，如“启动”“加速”等；车载雷达可探测较远处的固定路障；车轮上的速度传感器则负责监控车辆的运行速度；激光测距仪能够及时、精确地绘制出周边的 3D 地形图。智能车的感知机构为其决策机构提供数据支撑。

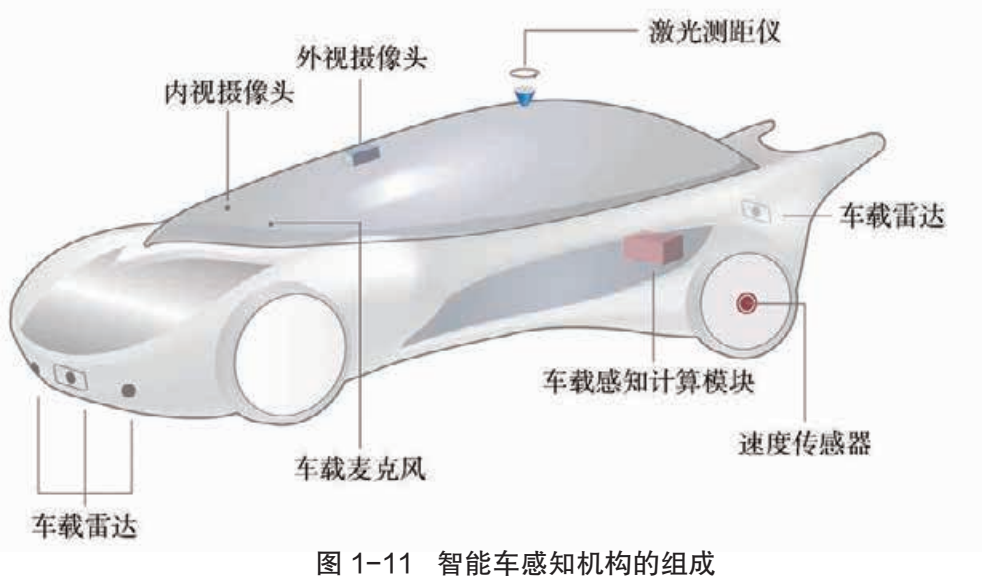


图 1–11　智能车感知机构的组成

（2）智能车的决策机构

人类大脑皮层约有 140 亿个神经元，掌控学习、记忆等活动。人的感知器官将收集到的信息通过神经系统传到大脑皮层，大脑根据这些信息解析出看到的物体、听到的声音、闻到的气味，进而对机体进行控制。大脑就是人类的决策中心，专门负责分析感知到的信息，控制机体进行反应。

同样，对于一个智能系统，决策机构是至关重要的部分，它掌控着整个系统的运转。决策机构不仅涉及诸如微处理器、计算机等硬件设备，还涉及处理、识别及控制等算法。

智能车的决策机构如图 1–12 所示。各类传感器像神经细胞一样与决策中心连接，收集到的信息或者发出的控制信号像电刺激一样在“神经元”间传输。如上所述，智能车的“眼睛”拍摄到的行车路况，或者速度传感器测到的实时行车速度，实质上是一组数字信号，它们经过预处理，以电平差的方式传输给决策中心。决策机构获取信号后进行处理与分析，识别出目标的类别和状态，根据目标点的坐标或者车辆期望速度，进行路径规划，选取控制策略。之后根据规划和策略，得到下一时刻的目标点坐标、车辆的期望速度、加速度等，向汽车电机以及转向轴输出控制信号，决定其前进、后退、旋转等。

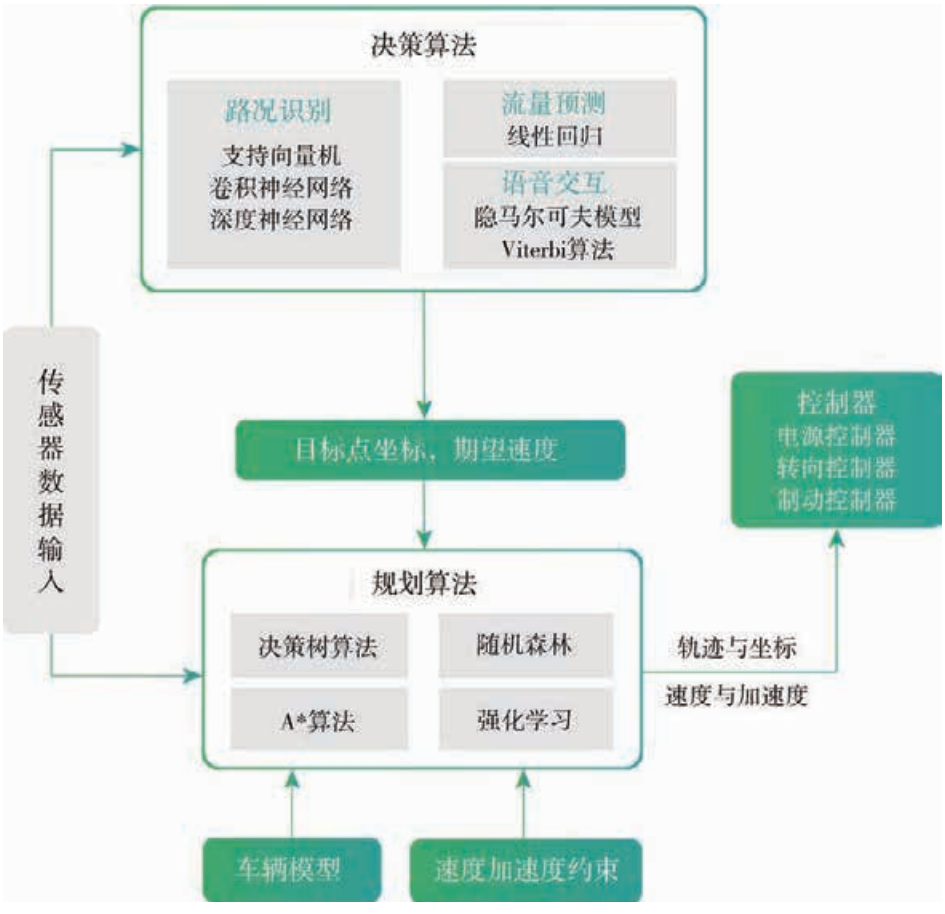


图 1–12　智能车决策机构结构示意图

（3）智能车的执行机构

我们的决策中枢——大脑，在分析感知信号之后，会得

到一个控制策略，例如行走、调节内分泌等。这些策略的实现，需要有力的执行机构。在人体中，大到关节、器官，小到细胞，都是可完成具体任务的执行器。执行机构使得大脑的决断变为现实，使智能体具有行动力。

智能车的执行机构具有同样的作用。电机、转向轴如同我们的关节。决策机构发出改变速度、角度的控制信号，然后由电机来改变速度，由转向轴来改变行驶方向。车内的空调、交互屏幕等类似于器官，控制改善着车辆的内部环境。从微观角度看，构成这些“关节”“器官”的机械零件或电子元件，如同细胞一样，通过精密配合构造出一个完整的执行机构。智能车的执行机构赋予了智能算法、智能策略可靠的执行力。有了执行机构的配合，算法才可能被实施。

（4）理解智能系统组成机构的关系

人工智能系统能稳定运行需要感知机构、决策机构和执行机构这三大机构相互配合。除此之外，还需要反馈机制的配合。

对于智能车来说，在控制车辆行驶速度或者旋转角度时，设定的目标速度或角度与实际行驶的状态是有差距的。因此，车辆通过采集速度及角度的传感器实时地对车辆的状态进行监测，实时计算当前运行状态与目标状态之间的差值，再利用这个差值对电机和转向轴进行控制，适当调整，减小差值，从而使车辆能够稳定运行。这种实时调整策略便是一种反馈机制（图 1-13），该机制保证了行驶的稳定性与流畅性。

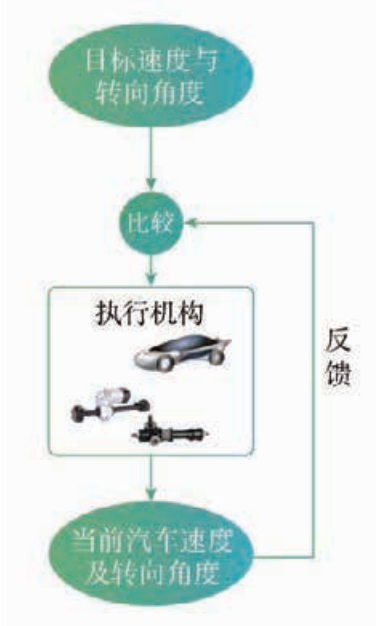


图 1-13　智能车反馈示意图

思考与讨论

理解智能系统的反馈机制，并使用自己的语言来描述智能车的反馈调节过程。

再如智能搜索引擎，当人们在搜索框中输入目标文字之后，搜索引擎终端会根据输入信息在互联网中进行检索比对，这个过程可以看作搜索引擎的感知过程。提取到初步的信息条文后，搜索引擎对其进一步筛选并且根据相关程度进行排序，最终确定在屏幕或者人机交互界面上呈现的信息，这一步为搜索引擎的决策过程；而搜索引擎在互联网中检索的过

程，在屏幕上呈现结果的过程，则是执行过程；用户根据呈现在屏幕上的搜索排序结果，决定是否点击某一个结果，先点哪个后点哪个，这是用户对搜索结果的反馈过程。系统可以根据反馈情况判断用户的偏好以及搜索条目排序的合理性。

可见，对典型的人工智能系统而言，感知机构是基础，决策机构是核心，执行机构是动力，反馈机制是保障。这四者相辅相成，才能构造出一个真正稳定的智能系统，从而保证智能系统能走进千家万户和工厂企业，为我们的生活带来便利，为生产带来效率。

小贴士

在实际应用中，人工智能系统不仅仅局限于智能车、机器人这类能够看得见摸得着的智能实体，还包括与工作生活息息相关的智能应用程序，例如网络搜索、语音助手等。

活　动

2.4 分组搜索相关资料，了解现在人工智能在各个领域的应用实例，举例分析智能系统的组成，用自己的话来解释感知机构、决策机构、执行机构与反馈机制之间的关系。

3. 明确人工智能的特征

在三大基本技术要素的支持、三大机构的配合以及反馈机制的保障下，当前的人工智能正在蓬勃发展。新时期的人工智能具备了以下特征：

（1）由人类设计，为人类服务，本质为计算，基础为数据。人工智能系统是由人类设计，以人为本的系统。这些系统按照人类预先设计的算法通过人类发明的硬件载体来计算或者工作。它们通过对数据的采集、处理、挖掘，形成有价值的信息和知识，为人类提供延伸人类能力的服务。

（2）能感知环境，能产生反应，能与人交互，能与人互补。人工智能系统应该能够借助多种传感器对外部环境进行感知，收集多种信息，同时能够对外界的输入做出不同的反应。此外，人工智能系统还需要具有通过一些外部设备与人类进行交互的能力，与人类合作，优势互补。

（3）有适应特性，有学习能力，能演化迭代，能连接扩展。人工智能系统要有一定的自适应能力和学习能力，即要有随任务变化、数据变化而自适应调节模型参数的能力。同时，能够与云端、客户端等实现数字化连接，不断进行演化迭代，提高系统的鲁棒性、稳定性、通用性。

小贴士

鲁棒性是 Robust 的音译，指系统的健壮性。

思考与讨论

根据自己对人工智能特征的理解，举例说明人工智能的特征。

知识链接

超级计算机

超级计算机的基本组成部件与个人计算机的组成部件无太大差异，但规格与性能则强大许多，具有很强的计算和处理数据的能力。其主要特点表现为高速度和大容量，配有多种外部和外围设备以及丰富的、高功能的软件系统。目前，“神威 · 太湖之光”超级计算机的计算速度最高可达到 12.5 亿亿次每秒，位于国际领先水平。超级计算机通常应用于天气预测、大气环流分析、天体物理模拟、密码分析等科学研究和国计民生问题。

与智能车相关的人工智能算法

智能车的主要功能模块包括物体识别、语音控制、人脸识别和自动决策等，如图 1-14 所示。

图 1-14　智能车的主要功能模块

物体识别模块在智能车中可实现对交通标志物、行人等目标的识别。它也可以应用在工业的残次品识别、医学的肺结节识别等情景。本书涉及物体识别模块的主要算法是神经网络和深度学习算法。

语音控制模块在智能车中可实现接收、处理和理解人发出的语音指令。它还可以应用在同声翻译、智能家居的控制等领域。本书涉及语音控制的算法是隐马尔可夫模型（HMM）算法。

人脸识别模块可实现智能车的刷脸开车门、乘客识别、安全认证等功能。它还可以应用在安防、金融等领域。例如，通过人脸识别来识别犯罪嫌疑人，或者在进行大额支付时，实现远程刷脸认证，以保证资金安全，防止资金被盗。本书在人脸识别模块主要介绍 k 最近邻算法（k-Nearest Neighbor，kNN）和支持向量机算法（Support Vector Machine，SVM）。深度学习等更先进的算法也已被应用于人脸识别。

自动决策模块在智能车中进行路径规划，找到最优路径。它也可以应用在商业的营销战略规划、游戏策略设计等领域。本书主要介绍决策树、A* 寻路等自动决策和搜索算法。

拓展阅读

人工智能的研究

人工智能与社会生产生活关系紧密，它的发展离不开各个方面的研究。人工智能的研究可以分为：基础层、技术层和应用层，如图 1-15 所示。

图 1-15　人工智能的研究层次

基础层：包括硬件 / 计算能力和大数据等研究。

技术层：包括赋予计算机感知分析能力的计算机视觉技术和语音技术、提供理解思考能力的自然语言处理技术、提供决策交互能力的规划决策系统、大数据分析技术，以及机器学习算法等研究。

应用层：包括金融、医疗、安防、交通、工业等行业的应用研究。

——摘自中国人民大学出版社《人工智能：国家人工智能战略行动抓手》

单元挑战　探究服务机器人

一、项目任务

服务机器人可以分为专业领域服务机器人和个人 / 家庭服务机器人。服务机器人的应用范围很广，主要从事维护保养、修理、运输、清洗、保安、救援、监护等工作。它由感知、决策、执行机构组成，并且要由算法、数据和计算能力这样的基本技术要素支撑。参考图 1–16 所示的家庭服务机器人，分组进行资料收集并讨论其三个基本技术要素的情况，然后对该机器人进行“拆解”，针对某一智能功能阐述它的主要机构组成。

图 1–16　家庭服务机器人

二、项目指引

1. 查阅资料，了解家庭服务机器人的各项指标，关注三个基本技术要素的信息。

2. 查阅资料，了解家庭服务机器人可以实现的功能，任选一个智能功能进行分析。关注服务机器人在运行某功能时调用的模块，理解这些模块的具体工作逻辑。根据具体工作逻辑将这些模块与主要组成机构相对应。

家庭服务机器人智能功能：_____

分析：

3. 整理资料内容，形成一份关于服务机器人的调查报告。或制作一个演示文稿，用于班级交流。

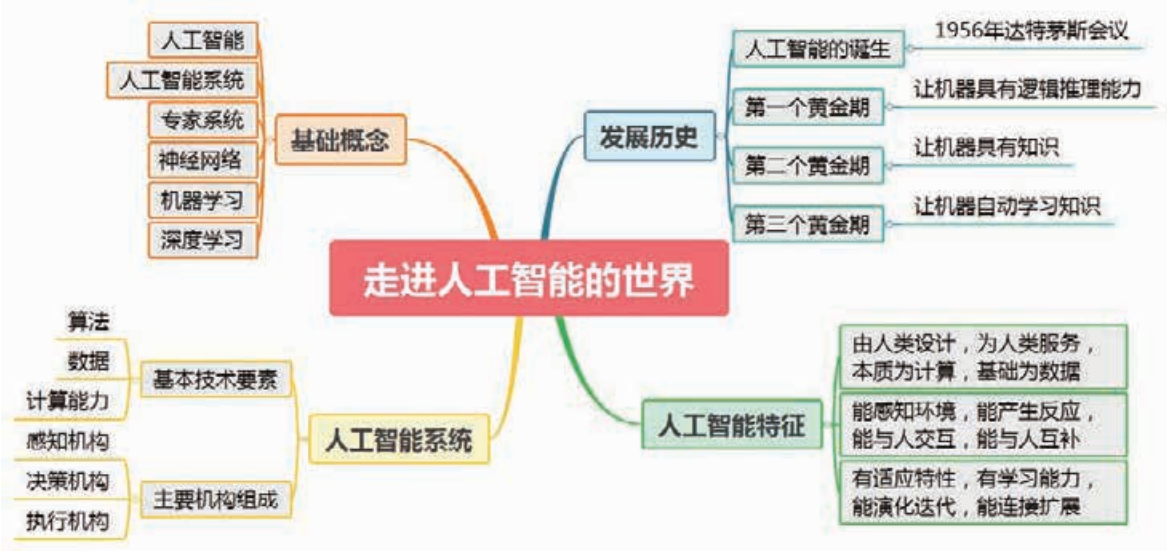
三、交流评价与展示

1. 每组展示交流各自选取的服务机器人的功能及分类结果。

2. 互相交流心得体会，加深对人工智能系统三个基本技术要素及主要组成机构的理解。

单元小结

一、主要内容梳理



二、单元评价

评价内容	达成情况
能用自己的语言阐述人工智能的概念（A、I）	
能列举应用了人工智能技术的系统（A、I、R）	
能说出人工智能的发展历史及各时期的特点（A、I、R）	
能说出 1956 年达特茅斯会议的意义（A、I、R）	
能以智能车为例描述人工智能系统的主要机构组成（A、I、R）	
能阐述人工智能系统的几个基本技术要素之间的关系（A、I、R）	
能用自己的语言阐述人工智能的特征（A、I、R）	
能说出神经网络、机器学习和深度学习之间的关系（A、I、R）	

说明：A—信息意识，T—计算思维，I—数字化学习与创新，R—信息社会责任

第二单元

理解人工智能技术的思想与方法

人工智能技术涉及面很广，主要包括感知、学习、推理、决策等方面。在实际应用中，人工智能最核心的一种能力是根据给定的输入作出判断或预测。当前，人工智能普遍利用计算机自我学习（机器学习）的方式来获得预测或判断的能力。机器学习已经成为基础性的人工智能技术，并已应用于计算机视觉、语音识别、自然语言处理等相关人工智能问题的求解。除了目前流行的机器学习算法，人工智能在几十年的发展过程中还积累了不少经典算法，虽然有些算法已经较少应用，但这些算法蕴含的算法思想至今仍然值得学习。

算法思想与应用方法构成了人工智能技术的精髓。本单元中，我们将通过剖析智能车的几种典型应用，如刷脸开门、道路流量预测、路况识别、路径决策、声音控制和车辆聚类等，学习人工智能的经典算法，了解人工智能的经典基础算法思想以及应用方法。

单元挑战

用 SVM 算法及深度学习给图像分类

学习目标

- 了解人工智能的核心算法。
- 了解人工智能技术应用的基本过程和原理。

项目二

让智能车能够“刷脸”开车门

——探究图像识别与理解

人脸识别技术是计算机视觉领域中图像识别的一个非常重要的研究方向。近些年，伴随着人工智能技术的进步和应用普及，人脸识别技术发展迅速，在金融、安防、商业等领域得到广泛应用。将人脸识别技术应用于智能车的设计，能实现“刷脸”开车门（图2-1）。这样的智能车可以没有车门把手，只需要车主脸部面对着车门的某个部位一照，车门便会随即打开；若是非车主，则车门不会开启。这既方便了车主，又提高了车辆的安全防盗性能。

图 2-1　用人工智能取代车门钥匙

项目学习目标

在本项目中，我们将探究把人脸识别技术应用于智能车的基本思想和方法。

完成本项目学习，须能回答以下问题：

1. 什么是图像识别技术？
2. 人脸识别是如何实现的？
3. 如何提取特征？
4. 人脸识别的常见算法有哪些？
5. 人脸识别的技术评价标准是什么？

项目学习指引

从万千人中识别出熟人，这对普通人来说是一件很容易的事情。人类判断“你是谁”，包括两个“处理步骤”：先找出人脸上具有区分度的特征；然后，根据观察到的人脸特征，把它和大脑里记忆中的某个人的特点匹配起来，得出“她或他是谁”的结论。这个过程，翻译成计算机的语言，其实就是“特征提取”与“分类器设计”两个步骤。“特征提取”即把对象示例的特征进行数据量化；“分类器设计”即产生一个从量化的特征数据到语义标签的映射关系。事实上，很多识别问题的解决都可以采用这两个步骤。

1. 了解图像识别与人脸识别技术

（1）图像识别技术

图像识别（image recognition）技术是指利用计算机对图像进行处理、分析和理解，以识别各种不同模式的目标和对象的技术。它一般利用数学模型，结合图像处理的技术来分析图像的底层特征和上层的语义信息，从而提取具有一定表达能力与区分能力的信息。

图像识别技术是“计算机视觉”研究领域的重要组成部分。生活中常见的图像识别应用有：人脸识别（face recognition）、表情识别、光学字符识别、手写体识别、医学图像分析、图像 / 视频内容检索等，如图 2-2 所示。

图 2-2　常见的图像识别应用

参见 P38 知识链接“图像识别”

（2）人脸识别技术

要实现人脸识别，一般要经过以下三个流程（图 2-3）：

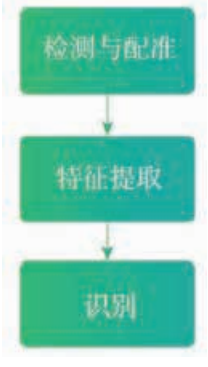


图 2-3　人脸识别的流程

① 检测与配准

基于人的面部特征，对输入的图像或者视频流，判断其中是否存在人脸，精确给出每张脸的位置框信息，以及五官、

小贴士

特征（feature）是人工智能中非常重要的概念。在计算机视觉领域，特征蕴含着图像中的核心信息。

轮廓的关键点位置信息。

② 特征提取

根据人脸框与五官点位置，结合图片本身，提取每个人脸中所蕴含的能表达身份特征的信息（如五官的尺寸、间距、纹理等），形成一个向量形式的人脸表征。

③ 识别

将需要被识别的人脸与已知的人脸数据库中的源人脸进行对比，计算与数据库中人脸的相似度，根据相似度分数确定该人脸的身份，人脸识别结果示例如图 2-4 所示。

图 2-4　人脸识别结果示例

活　动

3.1 体验人脸识别。

在网上找到可进行人脸识别的 AI 开放平台（图 2-5）。在此平台上找到人脸识别体验模块，体验人脸识别技术的多种应用，比如多人脸检测，思考其原理及日常应用。

图 2-5　某 AI 开放平台的人脸识别体验模块

2. 获取、表示人脸特征

要比较两张照片上人脸“长得像不像”，人是通过大脑的直观感觉来作出判断的，而人脸识别系统大多是通过将照片中人脸的有效特征（即可区分性特征）进行数字量化，然后分析代表这些人脸特征的数据值来进行身份判断。

特征的提取是指从目标对象中获取对于要做的任务有帮助的信息。特征提取是对图像理解和识别的关键步骤，是人脸识别技术的基础。

（1）提取人脸特征

人在探索世界时会通过耳朵、眼睛、鼻子等来获取外界的特征，形成听觉、视觉、嗅觉等感觉。神经系统将这些感觉转化成电信号传递给大脑。大脑会对比这些特征，并度量分析，从而得到外界的相关信息。为了让计算机完成特征提取，可以将这个过程抽象成具体步骤（图 2-6）：特征采集、特征使用、特征转换和特征度量。

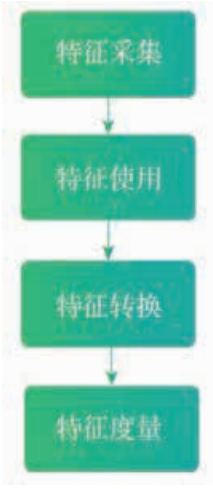


图 2-6　提取特征的流程

① 特征采集

特征的采集通常需要使用“传感器”。在人脸识别中，一般使用摄像头来获取人脸信息的全部特征，获得的特征以图像像素形式作电子存储。

思考与讨论

你认为还有什么方式能够采集人脸特征？

② 特征使用

经典寓言“盲人摸象”告诉我们，通常情况下，仅依靠象的牙齿、耳朵、腿、尾巴等一个个孤立的形状特征并不能进行智能分析。人脸识别同样如此。

以图 2-7 中的京剧脸谱为例，画师通过夸张的方式反映出某个人物角色的脸部特征，比如：关羽红脸、张飞面凶、曹操短须等。在进行脸谱判断时，如果只关注某一个特征，而忽略其他特征，往往会造成错误判断，如只看窦尔敦和张飞的下颌胡须，则很可能认为这是同一个角色。人脸识别系统也一样，需要对人脸的特征进行综合考察，才能得出最终的决策。如图 2-8 所示，目前的人脸识别系统，既会关注脸部的整体轮廓特征，也会关注五官的单个局部特征。


曹操　窦尔敦　张飞　关羽

图 2-7　京剧脸谱

图 2-8　人脸识别系统中的人脸特征

③ 特征转换与度量

获取的特征需要被转换成电信号人脑才能够处理，同样，“眼睛大”“皮肤白”等用自然语言描述的人脸特征也需要转换成计算机能够处理的数字形式特征。以对眼睛大小的描述为例，通常用自然语言可把眼睛分成“大”“中”“小”三类。若要将它们转换为计算机能处理的数字形式特征，可将“大”“中”“小”分别用数字 1、2、3 来表示（大小程度可以用整数，也可以用小数），从而形成一个计算机能处理的特征值。同样，其他特征也可以用数字来表示各自的属性类别或者等级程度。将这些特征值组合在一起写成向量的形式，就形成特征向量。有了特征向量，就可用它来描述物体的特征了。

有时候，等级不足以充分表示信息量，可选择用具体的数值直接作为特征值。以图 2-9 为例，假设收集到关于人脸的三个特征，分别是眼睛宽度 X_1、瞳距 X_2、鼻宽 X_3，它的特征向量可由如下方式形成：

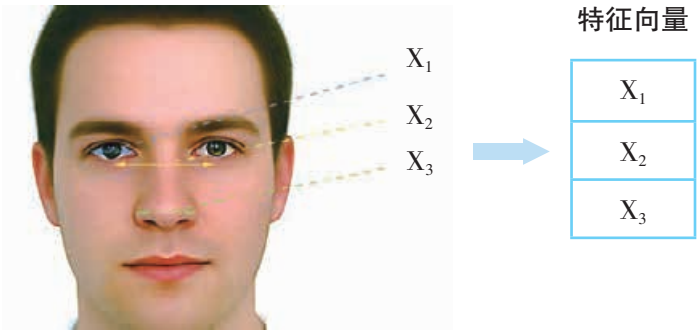


图 2-9　特征向量的形成

实际应用中特征向量维度通常不止三个维度，还可能包含表示其他内容的维度，以反映更多的特征信息。

小贴士

特征向量（feature vector）可将特征用数字化表示。二维、三维的特征向量与数学中的平面、空间坐标点的表示相似（高维的向量也可以类似地对等为高维空间的坐标点）。

通过这样的方式，将不同的特征组成为一个数字向量，可方便后续的计算机处理运算。这是一种常用的特征向量构建方法。

（2）用特征表示人脸

提取人脸特征之后，还需要根据应用需要，通过特征将人脸表示出来。

① 采用全局特征结构表示人脸

人们对人脸特征常这样描述：小张宽鼻梁，小李眼线比较长，小王下巴比较尖等。在人脸识别研究刚刚兴起的时候，科研人员关注的人脸几何特征也主要是眼、鼻、嘴等的形状以及它们之间的几何关系（如相互之间的距离、角度），然后将这些特征的数值与数据库中的人脸特征数值进行比对。

因特征数目少，该方法所占用的内存很小，且识别速度很快。但它其实是盲目机械地模仿人类直观的相互识别方法，简单浅显地分析人脸的部分表层结构特征，因此特征的表达能力有限，无法精确识别人脸（比如符合“宽鼻梁、长眼线、尖下巴”特征的人，可以找出成千上万），识别效果较差，不能将人脸所具备的如肤色、纹理、形状等丰富信息展现出来。目前，该方法已经很少应用。

为了获得全局的特征及结构，需先得到整张人脸的数据。首先需要将人脸的二维像素阵转化为一个一维向量。图像处理早期一般使用线扫（Zig-Zag）的方法来实现，图 2-10 展示了线扫的基本思想。

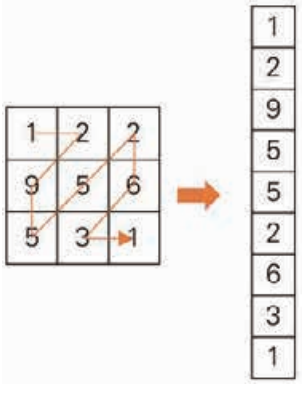


图 2-10　线扫方法

线扫方法的优点是方便简单，但是维度高，处理速度慢，容易被些许噪声以及几何变化如移动、旋转等所影响。为了免受这些影响，人们通过数据降维的数学处理方法，将原始图片像素进行变换，从而得到能够表征图片全局结构、形状、样式信息的特征。

数据降维的一种代表性方法是 PCA 算法。PCA 算法的原理类似“投影”。比如三维空间的一个球，往二维平面投影，则变成了圆。球是三维，圆是二维。在这个投影过程中，丢失了原来物体（球）的一部分特征（例如圆不再立体），但保留了原来物体的主要特征［例如圆（球）上每一点到圆心（球心）的距离相等］。

同理，人脸特征降维以后，人脸图像中的“噪声”、特征冗余和一些由细微的几何变化所导致的局部变化都被抑制掉了，最后得到的是能反映人脸全局结构、可用以区分的低

小贴士

PCA 算法，即主成分分析算法（Principal Component Analysis algorithm），它利用数学变换将人脸高维数据投影压缩到低维空间，并保留数据的主要特征（主成分）。

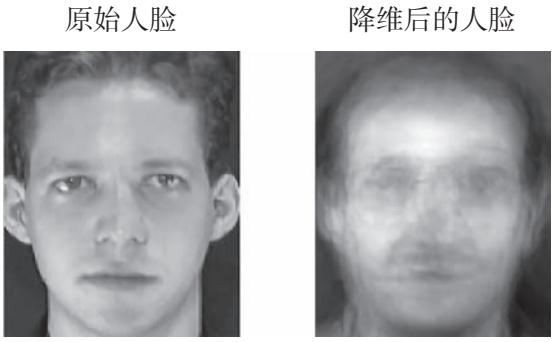


图 2-11　使用 PCA 算法对人脸数据降维

维度特征。这样既提高了人脸识别的计算速度，又降低了存储的复杂度。如图 2-11 所示，经 PCA 算法降维后图片虽然细节模糊，但完整地保留了人脸的结构。

② 采用局部特征表示人脸

很多时候，人脸的一个区分度非常高的局部特征能够帮助我们迅速、精确地识别这个人。在计算机视觉中，与全局特征对应的是局部特征。局部特征提取算法是相对精细的特征描述方法，它利用局部像素与整体像素之间的关系或者局部图像块之间的关系来获取某种变化信息。

LBP 算法（Local Binary Pattern algorithm）就是一种典型的局部特征算法。该算法通过比较中心像素和邻域像素的大小关系来得到人脸图像的角点和边缘等局部变化特征，然后根据这些局部变化特征区分人脸。

参见 P39 知识链接“图像像素”

最初的 LBP 算法定义在像素 3×3 邻域内，以该邻域中心像素为阈值，将周围 8 个像素的灰度值与其进行比较，若周围像素值小于中心像素值，则该像素点的位置被标记为 0，否则为 1。这样，3×3 邻域内的 8 个点经比较可产生一个 8 位二进制数（这里从左上角开始，按顺时针方向数），即得到该邻域中心像素点的 LBP 值，这个值可用来反映该区域的纹理信息。如图 2-12 和图 2-13 所示：

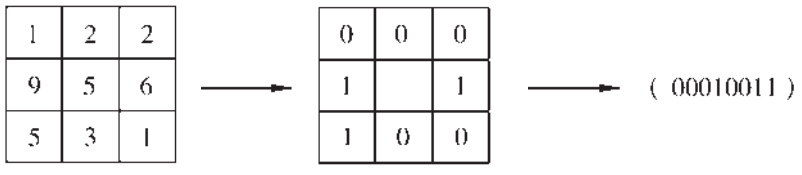


图 2-12　LBP 算法示例

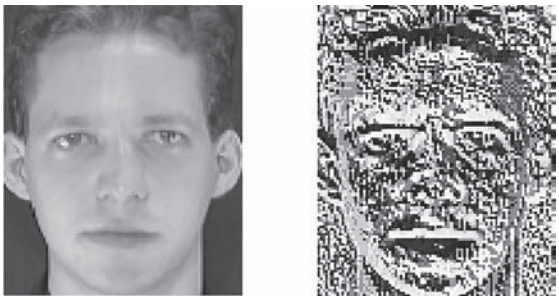


图 2-13　LBP 特征提取算法的输出效果

思考与讨论

经 LBP 算法处理后的结果有何特点？为什么它可以反映纹理信息？

LBP 算法具有灰度不变性、旋转不变性、计算速度快等突出的优点，因而在特征提取方面得到广泛应用。

通过以上几种方法对一张照片进行特征处理，就能得到该照片的一些关键信息点的数字信息（特征值）。人们通常将由一系列特征值组成的一个样本数据记为一个特征向量 $x=(x_1, x_2,\cdots, x_n)$。这些特征向量将是接下来进行分类、制定识别决策的依据。

3. 解析人脸识别原理

（1）分类问题

智能车的人脸自动识别解锁系统在获得了人脸的多个特征之后，需要对输入的特征进行处理，从而判断特征所属的人脸是不是车主的，并作出是否开启车门的决策。决策的内容实际上是将结果分成两类，即“是车主”和“不是车主”。这种智能决策问题被称为分类问题，这是机器学习领域非常重要的一个研究课题。分类的目标是判断一个新的样本属于数据库中的哪种已知样本类。

最普遍的分类问题是二分类问题，即类别数为 2（通常这两个类别分别用 1 和 0 表示）。当然，在很多时候，需要分类的类别不止两个，比如要开发一个可以容纳 4 位车主的人脸识别系统，那么需要分类的类别数就可能是 5（车主 A、车主 B、车主 C、车主 D、不是车主）。

通过这样的分类，智能车可以识别人脸是否属于车主，判断并决定是否允许对此人开启车门。

（2）人脸识别决策的方法之 kNN 算法

要利用提取的特征进行分类，可采用 k 最近邻分类算法。该方法的流程如图 2–14 所示。

简单地说，kNN 算法就是给出一个样本，计算在训练样本集中与该样本距离最近（最相似）的 k 个邻居，然后根据 k 个邻居所属的类别判定样本所属的类别。这里的距离具体

小贴士

判断前方是否有障碍物、预测某种产品市场价格的涨跌、判断某个病人是否得了某种疾病、分析某种药物是否有效等，均是典型的二分类问题。

参见 P39 知识链接“计算机视觉”

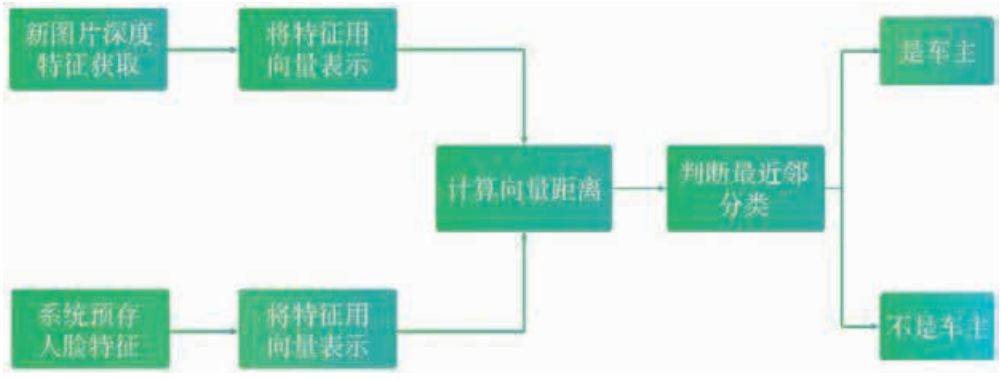


图 2-14　使用 kNN 算法进行识别的流程

到人脸识别中就是人脸的特征向量间的距离。

如图 2-15，在样本集中有“关羽”类（用  表示）和“曹操”类（用  表示）两个类别。X（用●表示）为待识别样本人脸，现要做的是判断该人脸是“关羽”类还是“曹操”类。按照 kNN 算法，假定 k 取值 5，然后计算出特征向量与之距离最近的 5 张人脸（图 2-15 中虚线所包围的人脸），从中可以发现，5 张人脸中“关羽”所占比例较大，那么这个未知人脸 X 将被归为“关羽”类。

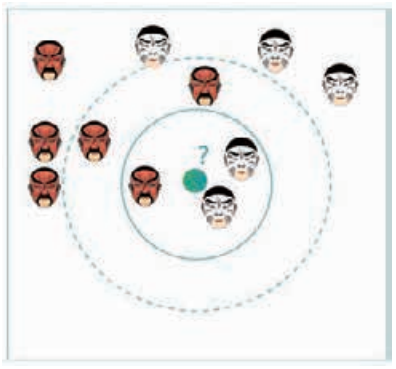


图 2-15　kNN 分类算法示意图

思考与讨论

上述案例中，如果 k 取值 3，那么分类结果会怎样？

kNN 算法的主要思想：先计算待分类样本与已知类别的训练样本集之间的距离（相似度），找到与待分类样本距离（相似度）最近的 k 个邻居；然后，根据这些邻居所属的类别，按照“少数服从多数”的原则，判断待分类样本的所属类别。主要步骤如图 2-16 所示。

图 2-16　kNN 算法主要步骤

根据 kNN 算法的思想，能很容易将一个新输入的特征归类到它最近邻的特征中，即找到最相似的人脸。如果新输入的人脸在数据库中并没有可以匹配的车主人脸，那么应将该特征归类到“不是车主”一类。智能车相应执行“不开车门”的命令。通常，我们通过对比特征距离是否超出同一个人的特征变动范围阈值（这个阈值通常是我们根据经验和实验尝试设定好的）来实施相关操作。对于超过了阈值的人脸，我们认为特征不匹配，即该人不是车主，不执行开门操作。对于没有超过阈值的人脸，并且通过 kNN 算法顺利归类，则认为特征匹配，即判断该人是车主，然后执行开门操作。

小贴士

当两个类别的数目一样多时，则给 X 随机指定一个类别。

活　动

3.2 假设数据库中有四张采集到的人脸的特征，四个特征向量分别为（0.1，2.8）、（0.4，3.1）、（2.5，0.9）、（2.8，0.7）（这里为了简化问题，假定每张人脸只有两个数字表示的特征），四张人脸的身份被认定为‘Guan’，‘Guan’，‘Zhang’，‘Zhang’，即前两张人脸特征被认为是‘Guan’（关羽），后两张人脸特征被认为是‘Zhang’（张飞）。

现在有一张新输入的人脸特征（0.7,2.6），大家试着运行配套资源中的代码，用 kNN 算法对之进行身份识别。

（3）人脸识别决策的方法之支持向量机算法

虽然 kNN 算法理论简单、有效，且易于实现，但它具有一些不可避免的缺点，如计算复杂度高，不同类别样本数量不均匀时预测偏差较大等。特别是在数据库庞大时，待识别的样本需要与数据库所有的样本进行比对，运算量非常大，以至于无法做到实时识别。

有一类算法，无论数据库多大，待识别的数据只需要被比对一次。这就是有参数分类器的算法，其中比较有代表性的是支持向量机（SVM）算法。

如图 2-17 中，关羽和曹操两类人脸的特征数据点（假设特征数据是二维的）分布于坐标系中，我们可以通过在坐标系中画出一条直线来区分这两类人脸的数据，并按此直线

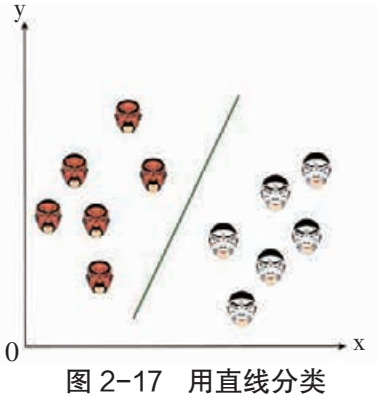


图 2-17　用直线分类

来对后续输入的人脸进行分类。因为在平面中，直线的函数表达式是：y=kx + b，其中 k 是斜率，b 是截距。如果能计算出比较合适的（k,b）组合，那么就可以使用这条直线对后续输入的样本进行分类。简单来说，就是当 kx+b 大于 0 时，可以判为正样本，反之，则判为负样本。

其实，满足分类要求的划分方法不止一种，也就是说这里可以有无数组的（k,b）组合，能产生无数条划分线。

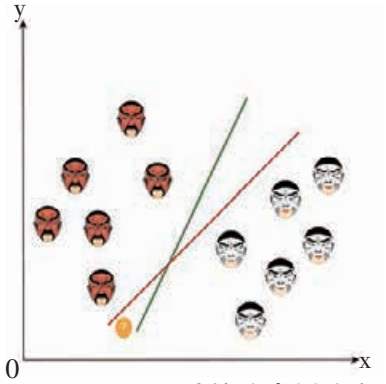


图 2-18　不止一种的分类划分方法

思考与讨论

如图 2-18 所示，红、绿两条直线都可以达到划分的效果，那该选择哪条划分线呢？你认为图中黄色问号处该属于哪种分类？理由是什么？

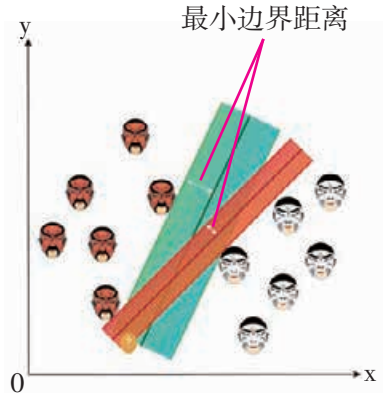


图 2-19　找最小边界距离中最大的分类器

使用 SVM 算法可以帮助我们选择合适的划分线。以图 2-19 为例，可以从图中直观地对比出红、绿两条划分线与正负样本点的最小边界距离。显然，绿色直线离最近的正负样本点的边界距离更大，它的两边所能留出的分类“缓冲”区域能够最大化。这样，即使出现一些模棱两可的样本，也能尽量以最大的概率留在“正确”的一边，增加了分类正确的概率。例如图中问号处为新输入数据，该数据处在模棱两可的边缘地带。如果按照绿色分类器（缓冲区域大），该数据为关羽；如果按照红色分类器（缓冲区域小），该数据为曹操。但是通过直观判断，这个数据显然更符合关羽（与关羽数据团簇距离更近），由此可认为绿色分类器更优。也就是说，最小边界距离越大，缓冲区域就越大，分类准确性也就越高。因此在实际分类时，需要找到最小边界距离最大的分类器。

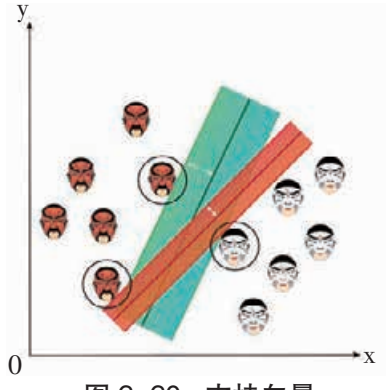


图 2-20　支持向量

要找到最大缓冲区域的边界，就一定要找到缓冲边界上的边界点。在这里，影响边界选择的几个数据点（数据向量），称为支持向量。简单地说，即这几个点影响了最终的边界距离，进一步决定了分类器的选择。这也是支持向量机算法的名称的由来。图 2-20 中，黑圈中的数据点即为支持向量。支持向量决定了支持向量机的参数以及分类性能。

4. 评价人脸识别性能

人脸识别性能在实际应用时可通过以下评价指标（表2-1）来衡量。人们可据此来对人脸识别方法作出最佳的选择。

表 2-1　评价人脸识别性能的标准

指标	说明
误识率 (False Accept Rate, FAR)	误识率是将其他人员误作指定人员的概率，例如把非车主误认为车主的概率。这个概率越低越好。从安全考虑，需要重点降低这个误识率，在系统设定中可以体现为降低判断阈值。误识率的计算公式为： $\mathrm{FAR}=\frac{\text{误识为车主次数}}{\text{车主识别总次数}}\times 100\%$
拒识率 (False Reject Rate, FRR)	拒识率是将指定人员误作其他人员的概率，例如把车主误认为非车主的概率。这个概率过高会影响使用人员的体验，因此需要尽量降低。拒识率的计算公式为： $\mathrm{FRR}=\frac{\text{误识为非车主次数}}{\text{车主识别总次数}}\times 100\%$
识别正确率 （Identification Rate）	识别正确率是正确识别人次与参与识别的总人次之比。这是一个整体的判断量，往往需要与误识率和拒识率综合考量。
识别速度	识别速度可理解为识别一幅人脸图像的时间或识别一个人的时间。这个时间越短，说明该系统的性能越好。然而，在其他因素不变的基础上，识别速度提高，往往会造成识别精度的降低。为此，需要更多地寻求硬件和算法上的改进。

在判断人脸识别的性能时，一般会结合以上四个指标进行综合考量。或根据需要，重点考量某个或某几个标准。如人脸识别技术用于私人场所门禁（车门、家门）时，误识率是一个更重要的标准。这时，误识率必须接近为0，否则将非主人识别为主人，会导致很严重的后果。而在另一种情况下，比如非关键通道的刷脸闸机（高铁入口），拒识率是一个更重要的量。因高铁闸口需要人流量快速通行，这时需要拒识率尽可能地低，否则大量乘客无法刷脸通过，不但影响乘客体验，也容易造成拥堵。

活　动

3.3 假定车主为关羽。如果让曹操进行五次车主身份人脸识别，结果如图 2-21 所示，请计算误识率。说说面对这个误识率，你会使用该产品吗？

图 2-21　活动 3.3 识别结果

3.4 假定车主为关羽。如果让关羽进行五次车主身份人脸识别，结果如图 2-22 所示，请计算拒识率。说说面对这个拒识率，你会使用该产品吗？

图 2-22　活动 3.4 识别结果

知识链接

图像识别

图像识别是指利用计算机对图像进行处理、分析和理解，以识别各种不同模式的目标和对象的技术。这是计算机视觉领域非常重要的一项技术。

图像识别技术是以图像的主要特征为基础。每个图像都有它的特征，如关羽脸谱图像的胡须长度、脸部颜色等。

图像识别大致需要经过四个步骤：①图像采集；②对图像预处理得到特征；③训练算法；④识别。

常见的图像识别应用有：物体识别、文本识别（图 2-23）、车牌识别（图 2-24）、人脸识别等。

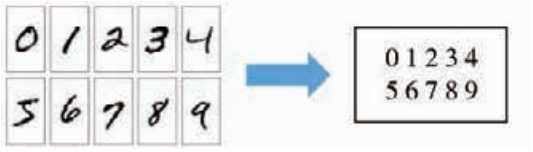


图 2-23　手写文本识别

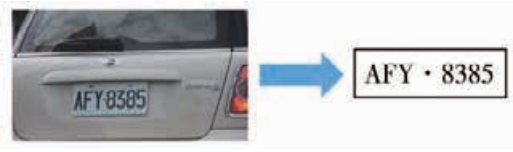


图 2-24　车牌识别

图像像素

如果大倍率放大计算机或手机上的图片，你会发现它们是由一个个细小的方块组成的。每一个小的方块就是一个像素（pixel），每个小方块都有一个明确的位置和被分配的色彩数值，所有小方块的颜色和位置决定了该图像所呈现出来的样子。

像素是组成图像的最小单位。每幅图像都包含了一定量的像素，这些像素决定了图像在屏幕上所呈现的大小。同样尺寸的图片，像素分布密集的图片会显得更加清晰，像素分布稀疏的图片会显得更加模糊，如图 2-25、图 2-26 所示。实际上，计算机对图像进行处理就是对图片中的一个个像素值构成的向量进行处理。

图 2-25　像素分布密集（高像素）的图片

图 2-26　像素分布稀疏（低像素）的图片

计算机视觉

计算机视觉是一门研究如何使机器“看”的科学，是使用计算机模仿人类视觉系统的科学。它的目标是让计算机拥有类似人类提取、处理、理解和分析图像及图像序列的能力。

在人工智能领域，计算机视觉主要研究用摄影机和计算机等代替人眼对目标进行识别、跟踪和测量等，并进一步进行图像处理，以模拟人眼的特性甚至延伸人眼的视觉范围，从而为人工智能系统获取信息。

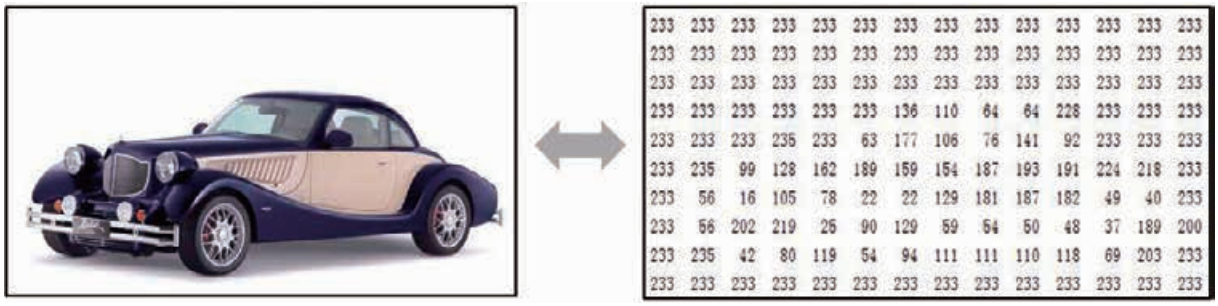
人工智能系统的计算机视觉包括两部分功能：

（1）模拟人眼——让机器去看。

（2）模拟大脑视觉皮层——让机器去理解。

计算机视觉在“看”的过程中有个很大的挑战，就是要跨越从最底层的像素值与高层

次的语义之间的“鸿沟”。摄影机不像人眼，它“看到”的并不是一个场景或者一个物体，而是一个完全由数字构成的像素表，如图 2-27 所示。



人眼看到的　　机器看到的

图 2-27　“看”的人机对比

除了“看”，计算机视觉的另一项任务是透过这一组数字，去挖掘出其中蕴含的语义信息。因此，计算机视觉中的各种算法和处理过程，本质上都是对数字的运算和处理。

目前，人工智能中计算机视觉领域的研究方向如图 2-28 所示。

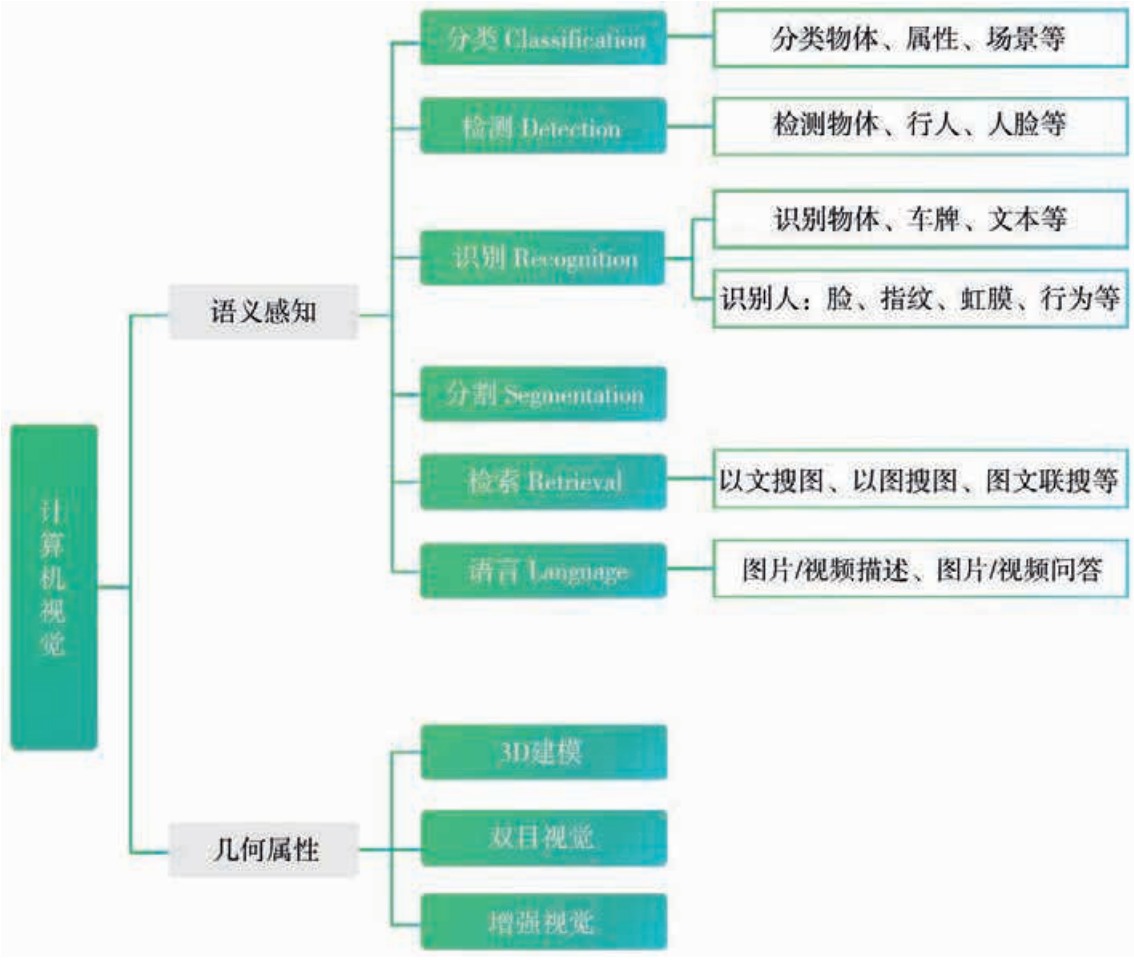


图 2-28　计算机视觉的研究方向

拓展阅读

图像分割

图像分割（segmentation）问题是计算机视觉领域中的一个重要课题，它是很多视觉研究内容的基础和关键步骤。

图 2-29　图像分割实例

图像分割指的是将数字图像细分为多个图像子区域（像素的集合）的过程，即把图像分成若干个特定的、具有独特性质的区域，并在其中提出感兴趣的目标的技术和过程。图 2-29 是一个篮球赛场景的图像分割实例。

从技术层面看，图像分割其实是一个对图像中的每个像素加标签的过程，它把属于同一区域的像素赋予相同的编号，便于后续的处理分析。例如可使具有相同标签的像素具有某种共同的视觉特性，比如同一颜色，以此来可视化分割结果。

图像分割可分为语义分割和实例分割。语义分割是根据图片内的语义内涵来对物体进行分割。实例分割则是根据场景中出现的不同的独立物体、事物来进行分割。

原图　　语义分割　　实例分割

图 2-30　图像分割对比

——参考 Ronghang Hu,et al.Learning to Segment Everything.2017

项目四

让智能车与用户对话

——探究语音交互技术

在我们的生活中，应用人工智能技术的产品已经非常多见，如各种智能音箱。通过与用户对话，智能音箱能按照用户的意愿播放音乐、新闻、故事等，它的背后是人工智能中语音交互技术的应用。设想如果将该技术应用于汽车，汽车也就能像人类的好伙伴一样跟人类进行交流（图 2-31）。用户只要通过语音即可命令汽车开窗、停车、开关空调以及播放音乐等。

图 2-31　让人工智能产品听懂语音

项目学习目标

在本项目中，我们将探究把语音交互技术应用于智能车的基本思想和方法。

完成本项目学习，须能回答以下问题：

1. 什么是语音识别？
2. 语音识别在生活中有哪些应用案例？
3. 隐马尔可夫模型对语音识别有什么作用？
4. Viterbi 算法的作用是什么？

项目学习指引

人机交互，简单来说，就是人与机器“沟通交流”。

从汽车诞生之日到汽车工业高度发达的今天，工程师一直在努力研究更好用的人与汽车的交互形式。人在开车的时候，手、眼都已经被占用，如果需要操作某些设备，如空调、雨刷器，语言交互显然有着比手动操作更方便也更安全的优势。现今，车载智能语音交互成为研究热点，并开始进入应用阶段。

活　动

4.1 了解当前车载语音交互的使用情况及人们的看法。

1. 认识语音交互

要实现智能设备与人的语音交流互动，必须要有语音交互技术的支持。它既可将人类的声音变换为计算机可以理解的信息，也可令计算机生成自然语言，并通过语音的形式对人类进行回应。

（1）语音交互系统中的主要模块

语音交互系统（图 2-32）通常涉及多个模块，如语音的识别与理解［语音识别（Automatic Speech Recognition，ASR）和自然语言理解（Natural Language Understanding，NLU）］模块、对话决策模块（参考历史输入信息），以及最终的自然语言生成（Natural Language Generation，NLG）和语音合成（Speech Synthesis Technology，SST）模块等，详见表 2-2。

小贴士

“交互”的意思是交流互动。语音交互指的是人类和机器通过语音交流互动的方式来实现某种功能。

图 2-32　语音交互系统

表 2-2　语音交互系统中的主要技术模块

模块	说明
语音识别	语音识别也称为语音转文本识别（Speech To Text，STT），它负责将收集到的语音信息转换为对应的文本信息，为机器进行下一步的自然语言理解作准备。这是机器具备“听懂”人说话的能力的基础。 语音识别技术被广泛应用在车载导航、外语教育和需人声验证的场合。
自然语言理解	自然语言理解可使机器理解文本，从而可理解语音识别模块输出的文本的含义，并确认需要完成的任务。 自然语言理解技术被广泛应用在搜索引擎、输入法中。
对话决策	对话决策的任务是使机器完成语音指定的功能，或者确定与用户下一步的交互内容。一般来说，对话决策需要借鉴历史输入信息来完成判断，以使机器能更精准地与用户进行交互并且完成任务。 对话决策模块是语音交互系统的中心决策模块。
自然语言生成	自然语言生成可以看作是自然语言理解的逆过程，它负责把要阐述的概念以一定的语义和语法规则生成一段自然语言文本。 自然语言生成技术被广泛地应用在人机对话系统、新闻内容生成等方面。
语音合成	语音合成又称为文本转语音（Text To Speech，TTS），它可根据文本生成对应的人声语音，使机器具备模仿人“说话”的能力。 语音合成技术已经被广泛应用在地图导航、语言翻译等方面。

图 2-33　人与智能车的对话情景模拟

（2）语音交互过程

若要实现如图 2-33 所示的人与智能车的对话情景，智能车系统大致需要经历如下的过程。

首先，利用语音识别模块将人发出的声音变换为文本，接着通过自然语言理解模块确定语音指令内容。比如人的第一句语音指令内容“智能车，播放音乐！”可被理解为“播放音乐：音乐题目未指定，音乐风格未指定，歌手未指定……”

然后，根据语音指令内容，对话决策执行相应操作。比如对话决策会向用户征询是否播放其最常听的音乐类型。这个过程中，为了利用语音同人进行交流互动，系统会生成自然语言，并且模拟人声播放出来。应注意到，对话是一个连续的交互过程，因此，有时候需要借鉴历史输入信息进行判断。比如智能车在得到指令“换一首古典的吧！”之后，会

根据历史输入内容“播放音乐”将指令内容更新为“播放音乐：音乐类型指定为古典”。

语音交互中的这几种技术在机器和人之间的交互中发挥着重大的作用。这些技术如今被广泛应用在我们的生活中，并逐渐改变着我们的生活方式。

参见 P50 知识链接“语音合成”“自然语言理解与生成”

思考与讨论

语音交互技术如何改变人们的日常生活？它主要应用于哪些方面？

没有自然语言理解，机器能真正实现语音识别吗？为什么？

活　动

4.2 在网上找到可进行语音识别的 AI 开放平台（图 2-34），体验将语音变成文本的语音识别和将文本变为声音的语音合成效果。

图 2-34　某 AI 开放平台的语音识别体验模块

2. 让机器理解语音

语音交互技术中，自然语言理解环节非常关键。只有准确理解了语音内容，才能进行之后的决策和输出。

人类听到某句话之后，可能会根据往日的经验，在脑海中把这句话切分成一个一个的单字，然后根据这些字的读音，按先后顺序恢复出对应的汉字。

这看起来是一个简单的声音到文本的变换过程，但在该过程中存在一个难点：在汉字中，每一个字音对应的汉字可能不止一个（即同音字）。如发音“yuè”的字可能是“乐”“越”“月”等。对于人类而言，仅凭单字音也很难确定文字，但如果给出这个字前后的一个字或者几个字，人类便可以根据上下文的含义确定该字，如在前面给出了“yīn”，“yuè”的候选范围就大大缩小，最终可确定为“（音）乐”。

如今，人们使用隐马尔可夫模型（HMM）来让机器模拟人类这样的智能过程。

当一个人想说某句话时，会首先在脑海中形成这句话的每一个字，然后通过声音传达这句话。听者在听到这句话的时候，是不能直接观测到说话者脑海中的文本信息（隐含节点）的，但是可以听到说话人的声音（可见节点），并且将每一小段声音同文本对应起来。听者通过猜想所有可能的文本，再将其与听到的上下文信息进行比对，选取最有可能的文本序列。

隐马尔可夫模型正是站在人类听者的角度对文字进行识别。如图 2-35 所示，隐含节点按照字的先后顺序链接成单向链，并且每个隐含节点产生一个可见节点。“播”“放”“音”“乐”四个字是 HMM 中的隐含节点，而由这些字生成的语音片段就是可见节点。

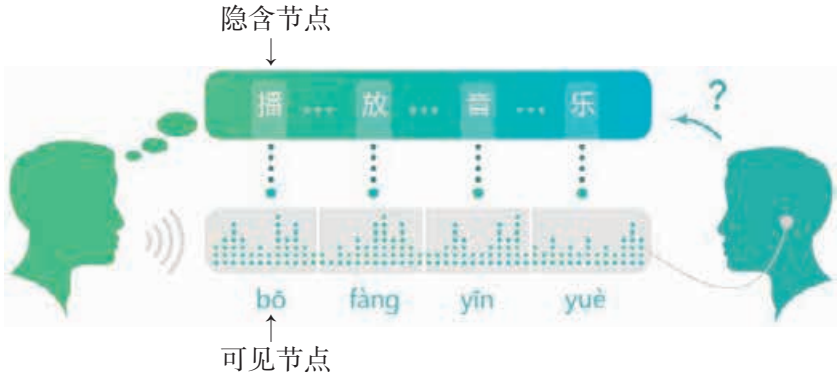


图 2-35　隐马尔可夫模型示例

HMM 的特点在于每一个隐含节点只与前一个隐含节点以及当前的可见节点输出相连。比如例子中的“乐”这一隐含节点只与“音”隐含节点以及对应的声音输出片段“yuè”有关，与“播”“放”等隐含节点以及“bō”“fàng”“yīn”等可见节点无关。这样的设计，使得在计算隐含节点的概率时，可以从当前发音和上文内容推算出对应当前隐含节点的可能性最大的文本。

小贴士

马尔可夫模型（Markov Model，MM）指的是一条连续的状态链，其中状态和状态之间可以相互转化并且每一个状态都只由前一个状态转移得到。

隐马尔可夫模型（Hidden Markov Model，HMM）中的“隐”的意思是它的状态不可被直接观测到。

小贴士

在实际的 HMM 算法应用中，隐含节点代表的往往是音素（比单字音更小的语音单位）的不同状态。另外，语音片段在输入到可见节点时，还需要对其进行特征提取等预处理操作（参见 P51 拓展阅读“语音识别特征——梅尔频率倒谱系数”）。

活　动

4.3 尝试在图 2-36 中从左到右连接所有可能的词组（如图中的“歌曲”和“很懂”），再选出最有可能的句子来。体验了解隐马尔可夫模型的识别过程。

图 2-36　活动 4.3 配图

3. 用算法实现语音识别

在语音识别中，HMM 的作用是找到可能性最高的隐含节点序列。HMM 可以极大地减小目标隐含节点的搜索范围，这得益于 Viterbi 搜索算法。

中文每一个字音都可能对应很多候选字，因此，所有字的组合结果数量极其庞大。如“古典音乐”，单纯从字音对应过来的组合有（古、鼓、谷、股、骨……）+（点、电、典、碘、店……）+（音、因、茵……）+（越、月、乐、悦……）。若使用穷举的暴力搜索方式会产生巨大消耗，导致问题的难度大大提升。Viterbi 算法正是为解决这一点而提出来的。

如图 2-37 所示，Viterbi 算法可计算一个隐含节点（字）转移到另一个隐含节点（字）的可能性大小，并且剔除可能性（概率）小的枝节，使它们不会被继续搜索。例如，对于“典”字这一节点的状态而言，“谷典”这个词出现的概率（或者说出现从“谷”到“典”这一组合的可能性）是非常低的，因此，“典”这一节点会剔除掉从“谷”而来的隐含节点路径，只选取概率比较大的“古”这一路径（图中以粗点线标注）。对该层每一个字进行这样的操作，在猜测每一个字时只保留概率最高的路径，一直进行到最后一层，将可能性最高的序列作为搜索结果输出。

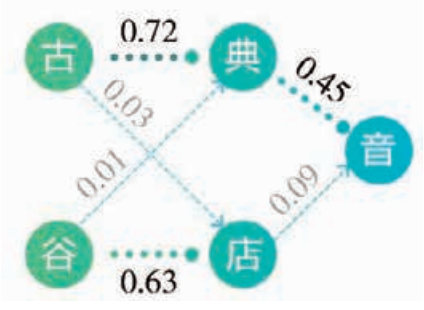


图 2-37　简单的两层 Viterbi 的例子

思考与讨论

Viterbi 算法还可以用在哪些搜索情景中？

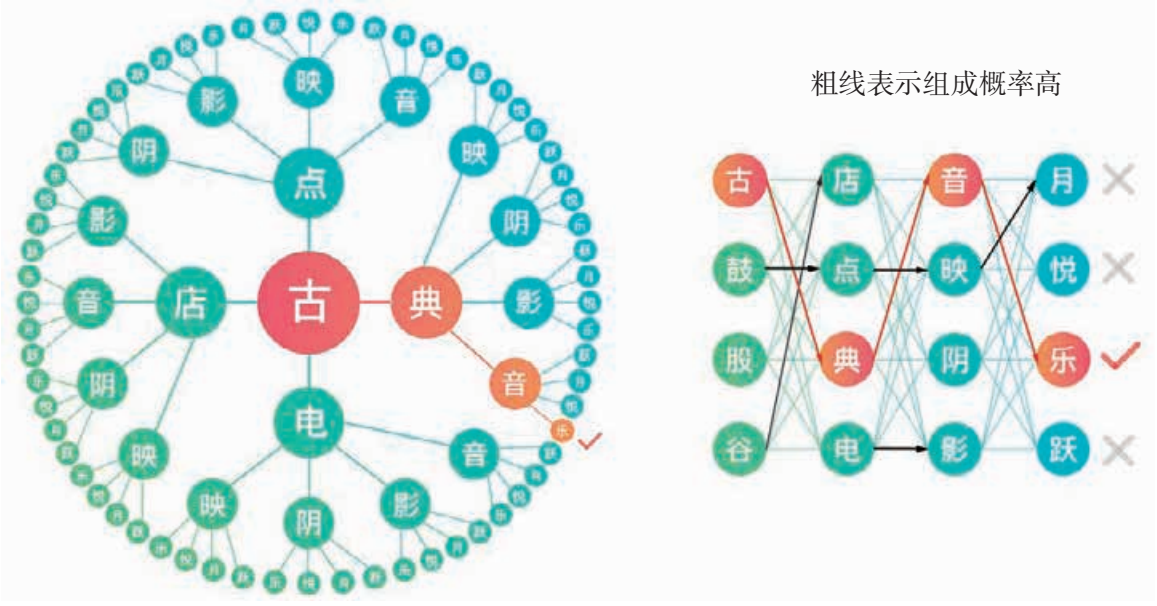


图 2-38　穷举（左）和 Viterbi 算法（右）在四个隐含节点下的计算量

从图 2-38 还可以看到，“音”这一节点有来自上一节点的“典”和“电”两条概率比较高的路线（不一定要组成词，也可能是经常一起按顺序出现的字），但是由于“电”之前的候选字没有和“电”组成概率比较高的路径，因此，从“电”传来的路径的概率比从“典”传递来的要低。这是该模型后续能正确选择下一条路径的基础。另外，从图中可看到，会出现两条概率相对较高的路线——“古典音乐”和“鼓点映月”。尽管后者在语义上说不通，但由于隐马尔可夫模型仅关注相邻字之间的转移可能性，后者仍有非常高的概率，这对隐马尔可夫模型语音识别的鲁棒性和准确率造成不利影响。在实际操作中，后一情况可以通过更长的隐含节点链以及更精准的字音（比如音调、翘舌音和前后鼻音）推断来避免。为了对每个字的上下文基于更多的历史数据进行分析，现在通常使用循环神经网络（Recurrent Neural Network，RNN）及其改进算法“长短期记忆网络”（Long Short Term Memory，LSTM）来提高识别准确率。RNN 和 LSTM 能够比隐马尔可夫模型回顾更长的时间序列信息。

参见 P50 知识链接“循环神经网络（RNN）”

4. 应用语音交互技术

基于 Viterbi 搜索算法的隐马尔可夫模型在语音识别中发挥着核心作用。语音识别作为语音交互的重要一环，其识别准确性的提升使得语音交互技术在日常生活中大有可为。如

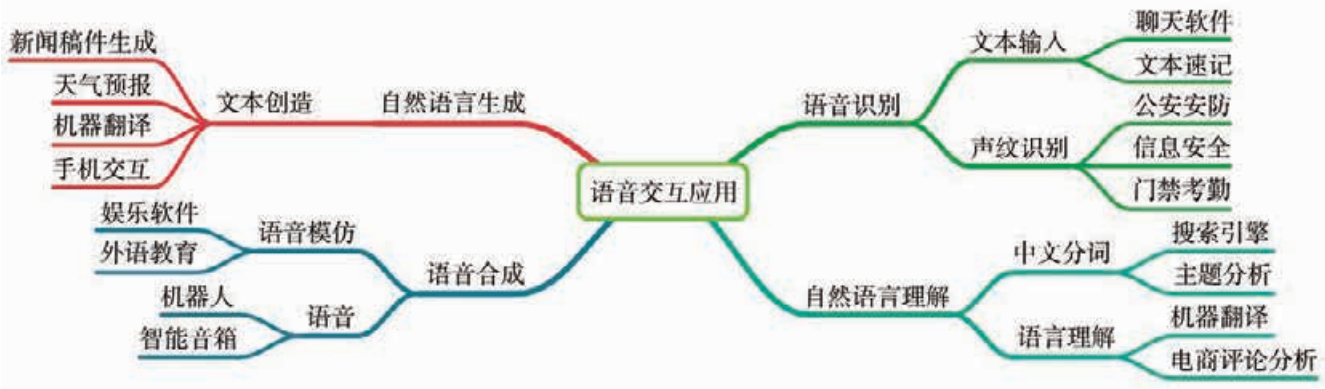


图 2-39　语音交互应用一览

图 2-39 所示，语音识别技术不仅可以作为一种辅助的文本输入方式，还可以用在声纹识别等安全防护领域。除此之外，其他语音交互技术在生活中也应用广泛，如自然语言理解技术用在机器翻译、评论分析等情境中，语音合成技用于自动合成外语、音乐等，自然语言生成技术用于生成天气预报和新闻稿件等，这些应用不仅方便了人类生活，也节省了大量人力资源成本。

语音交互技术属于计算机科学领域与人工智能领域的一个研究方向，它仍然存在很多问题亟待有效解决，比如短语音的有效识别、识别系统对口音的适应性以及在复杂环境下的抗干扰语音识别等。

另外，语音交互常用于安防，这让人们关注到语音攻击（比如重播录制的目标声音以欺骗语音识别系统，或超声波干扰语音识别），并对其展开防卫研究。

活　动

4.4 尝试进行简单的语音识别分类。打开配套资源，按照语音识别的步骤提示准备数据并运行代码，查看模型的分类准确率。

注意：可自己录制简单的音频样本进行训练和测试。数据应存储于 Python 代码同级目录下的“train”和“test”文件夹（分别存放训练和测试用的音频文件）。音频文件名格式为“类别号 _ 序号 .wav”。

4.5 以小组为单位探讨为智能车设计语音交互系统的技术方法，包括该方法的工作流程、各流程环节可能用到的技术及技术原理，在班级中交流介绍。

知识链接

语音合成

语音合成是语音交互系统的“嘴巴”，它负责将机器生成的文本通过声音的形式表达出来。语音合成的评价指标包括：语句意思是否被准确表达，生成的语音是否接近人声。前一个评价指标衡量声音是否能被人听懂，而后一个指标衡量生成的声音与人声的相似度。

语音合成最简单的方法是根据文本找到预先录好的对应语音片段，再将这些语音片段串起来。这种做法虽然简单，但是发音效果不够流畅，发音相对比较生硬，与人的自然发音存在明显的差距。随着大数据时代的到来，当前主流的语音合成方法是基于深度学习的语音合成。这样的语音合成可学习更多的历史数据，从而使得发音更加准确、自然。

语音合成技术现已得到广泛应用，给人类带来很多帮助。比如结合了语音合成技术的智能手机，让盲人使用手机成为可能。语音合成技术还可帮助发音障碍人士与他人进行交流，比如丧失语言能力的霍金曾依靠语音合成输出设备与世界各地的人们顺畅沟通。

自然语言理解与生成

自然语言理解与生成是语音交互系统的“翻译器”，它负责使计算机能够理解并使用人的语言。自然语言理解通常使用编码器实现“人类语言”到“机器编码”的转换，反之，自然语言生成则使用解码器实现“机器编码”到“人类语言”的翻译，如图 2-40 所示。

图 2-40　人类语言与机器编码的转换

目前自然语言理解与生成技术被广泛应用在文本翻译以及搜索引擎中。该技术的引入使得翻译结果更加“人性化”，也使得搜索结果能够更加精确地指导用户找到信息。相比于传统技术，该技术能够使机器更智能地理解人类的需求，与人类进行更加高级的交互。

循环神经网络（RNN）

在语音识别中，分帧单独运用 HMM 的方法可以获得一定的识别准确率，但 HMM 对于一个字只与相邻字相关的假设影响了其识别准确率。改进的方法是对连续多个字之间的关联性进行关系建模，并且从前向和反向（比如前文中提到的从“古”到“典”以及反向

从“典”到“古”）同时分析。基于这一目的，RNN 作为 HMM 的辅助被提出作为语音识别的模型。

RNN 结构简单（图 2-41 中，A 表示相同的 RNN 处理单元），其工作原理是对时序的每一次输入都使用相同的迭代单元结构进行建模。A 接受每一小段语音的特征，输出结合历史信息后的 RNN 特征，并将“记忆”传递到下一阶段。RNN 相对于单独使用 HMM 而言，可对时间间隔更长的先前节点状态进行分析建模，从而避免犯“鼓点映月”这种错误。

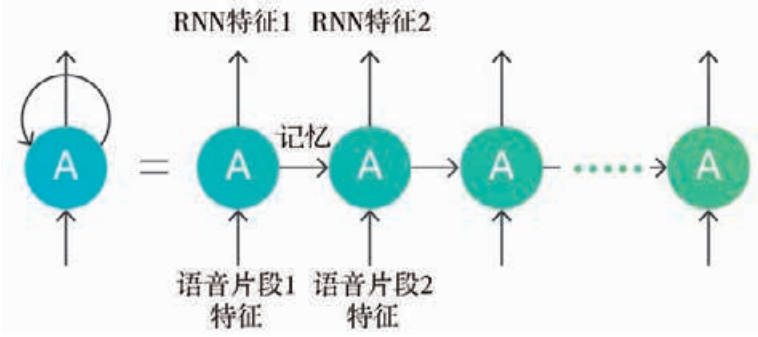


图 2-41　RNN 结构

拓展阅读

语音识别特征——梅尔频率倒谱系数

与人类不同，机器在执行识别任务时，不是直接进行识别，而是先对输入信息进行特征提取。语音识别特征提取的作用是滤除掉一些与语音识别任务无关的干扰或者噪声，从而使机器能够更“专心”地训练语音识别技能。常见的语音识别特征是梅尔频率倒谱系数（Mel-Frequency Cepstral Coefficients，MFCC）。梅尔频率倒谱系数利用人耳对不同频率声音（音调）的敏感程度不同，将声音的音调信息通过非线性的方式映射到梅尔频谱中。在该频谱中，人所能感知的音调变化是线性的，比如该频谱上差两倍，人类感觉上的音调变化也是两倍。

梅尔频率倒谱系数的产生过程是：将音频数据分帧，采集其频率数据，再用滤波器得到梅尔频率数据，最后得到 MFCC 系数。该特征提取算法的前提假设较少，适用情景非常广阔。梅尔倒谱系数经试验证明有着相对较好的识别性能，因此作为一种语音识别特征被广泛使用。

图 2-42　梅尔频率倒谱系数的产生过程

——摘译自 Xuedong Huang,et al.Spoken Language Processing: A Guide to Theory,Algorithm and System Development .2001

项目五

让智能车自动规划路径

——探究智能决策与搜索算法

智能车要实现无人驾驶，必须要能够自动规划路径，即能根据起止地点，自动寻找最佳路径，并行驶到达。路径规划技术在我们日常生活中的很多领域已经得到广泛应用，比如手机地图导航软件里的路径规划（图 2-43）、扫地机器人的清扫路线规划、电子游戏中游戏角色的移动路线规划等。

图 2-43　地图导航软件的路径规划

项目学习目标

在本项目中，我们将探究把自动路径规划技术应用于智能车的基本思想和方法。

完成本项目学习，须能回答以下问题：

1. 要实现自动路径规划需要解决哪些问题？
2. 什么是时间序列？
3. 什么是决策树？
4. 启发式搜索的基本思想是什么？

项目学习指引

自动路径规划是实现汽车智能化的关键技术之一，其主要任务是依据环境感知系统处理后的环境信号以及相应的地图信息，在满足汽车行驶诸多约束的前提下，以某种性能指标（如最少时间、最短距离）最优为目的，规划出车辆的运动路径。

在车辆实际行驶的过程中，交通路网状态会随时间变化而变化，其中包含了静态交通限制信息（如短期施工）和动态交通流量信息（如交通流量状况、各路段限速状态），这些都是不确定因素，需要采集实时信息。在智能车的路径规划研究中，核心是路径规划的算法，算法的选择将直接影响到路径规划质量的优劣。

1. 用人工智能实现路径规划

路径指连接起点位置和终点位置的序列点或曲线，而构成路径的策略称为路径规划（path planning）。在人工智能领域，路径规划是让智能设备具有自动规划路径的能力的技术。

通常路径规划的流程为：先通过交通流量预测判断道路的拥堵情况，再结合其他因素进行智能决策以排除交通状况不佳的路段，最后采用搜索算法找到一条路程最短的路线。

小贴士

交通流量预测、智能决策和搜索算法是路径规划中的三个重要部分。

图 2-44　路径规划的流程

如图 2-44 所示，从起点到终点至少有四条路线可以选择（分别用红、绿、黄、蓝四种颜色表示），但其中有一条路线（绿线）正在施工。交通流量预测模块首先预测每一个路段的车流量，车流量过大的路段被认为可能会拥堵（如蓝线所示路线）。接着智能决策模块会将难以行驶的施工路段（绿线）、拥堵路段（蓝线）筛选掉。随后搜索模块从剩下可以通行的两条路线（红线和黄线）中选出路程最短的一条。

活　动

5.1 选择一种在线地图，搜索本市（县）地图。在地图中规划从家到学校的路线，比较不同方式（步行、骑车、公交、自驾等）及不同时间点地图给出的不同信息，思考其原因及路径规划的侧重点。

2. 预测交通流量

通常，路口的车流量并不是恒定不变的。交通流量预测是利用时间序列预测的一种应用，即通过统计路口的历史车流量数据和跟流量相关的一些特征数据来进行交通流量预测。

（1）用线性回归方法预测

交通流量预测可采用线性回归方法。

回归，即是“由果索因”。这是一种归纳的思想，即根据大量数据所呈现的状态，推断出数据之间蕴含的数学关系。例如在地球表面上抛掷一个苹果，它的高度与运动时间总是呈现二次函数的关系。回归有两类，当这个数学关系为线性关系时称之为线性回归，否则称之为非线性回归。回归方法主要解决如何通过样本来获取最佳的拟合线。

小贴士

时间序列指将同一统计指标的数值按其发生的时间先后顺序排列而成的数列。如图 2-45 所示，车流量可以看成一组按时间先后顺序排列的数列，即时间序列。

线性回归（linear regression）是一种应用较为广泛的回归算法，它可以用来预测或者分类，主要是解决线性问题。非线性回归（nonlinear regression）也可以通过某种分析方式转化为线性回归。

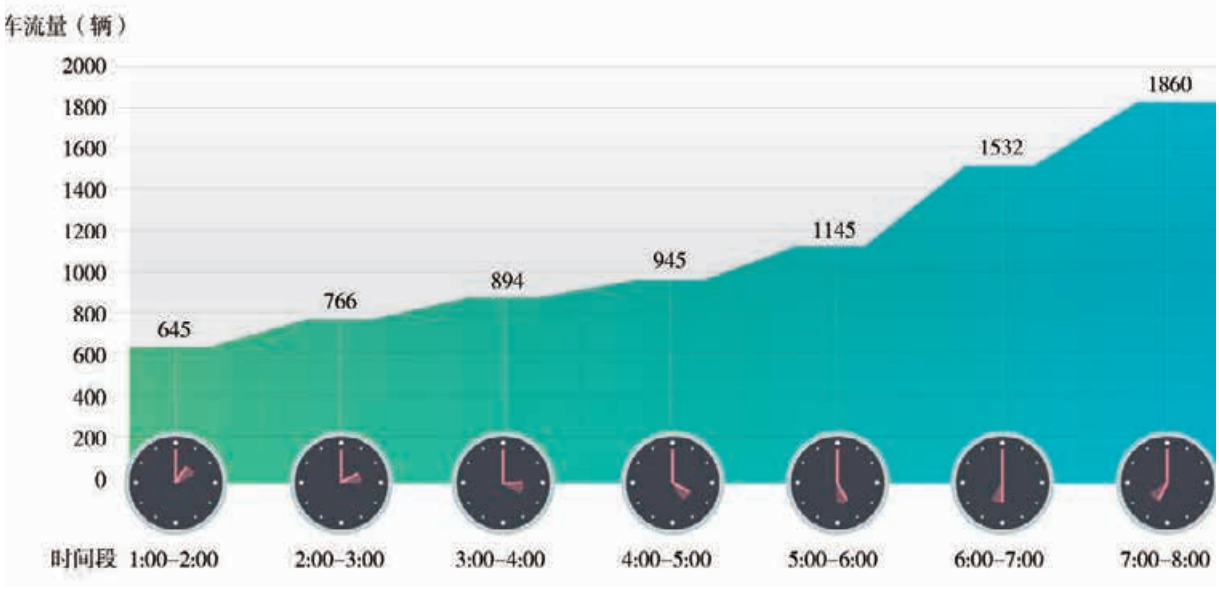


图 2-45　某路口某日 1 时至 8 时的车流量统计图

如果把车流量的预测抽象成线性回归问题，通过对时间（用自变量 x 表示）和车流量（用因变量 y 表示）进行建模，则可

建立车流量 y 与时间 x 的关系函数 y=f(x)。像这种自变量只有一个的线性回归被称为单变量线性回归。若是多个自变量的线性回归则称为多变量线性回归。本项目假定车流量只和时间有关，因此路口车流量预测是一个单变量线性回归问题。

以表 2-3 为例，对于表中的变量关系，可假设关系函数为 y=ax+b，当 a=2，b=2 时，这个假设的函数即为 y=2x+2。如果输入 x 为 1，输出的预测值 y 为 4，与右表中 y 值（7）的差值为 3。这说明这个由 a=2，b=2 构建的关系函数不是特别理想。所以我们应努力调整参数 a 和 b 的值，让函数的输出预测值与真实值差距最小。

参见 P63 知识链接“多变量线性回归及梯度下降”

表 2-3　单变量线性回归示例

x（自变量）	y（因变量）
0.5	4
1	7
2	7.5
3	8

（2）优化函数的方法

人们一般通过构建代价函数（cost function）来衡量预测值与真实值的差距。这里采用一种常用方法：最小二乘法。

最小二乘法（least-squares method，又称最小平方法）是一种数学优化技术。它通过使误差（预测值与真实值的差值）的平方和最小化来寻找数据和函数的最佳匹配。利用最小二乘法可以简便地求得未知的数据，并使得这些求得的数据与实际数据之间误差的平方和为最小。

从公式角度理解，就是要在给定数据对 $(x^{(1)}, y^{(1)})$，$(x^{(2)}, y^{(2)})\cdots(x^{(m)}, y^{(m)})$ 及假设函数 y=ax+b 的条件下，求得 a 和 b 的值，使得代价函数的值 $J(a, b)=(y^{(1)}-ax^{(1)}-b)^2+(y^{(2)}-ax^{(2)}-b)^2+\cdots+(y^{(m)}-ax^{(m)}-b)^2$ 达到最小。将上例表中的数据代入，可求得 a=1.339，b=4.449，故上例使用最小二乘法回归得出的自变量 x 与因变量 y 的关系为：y=1.339x+4.449。

如图 2-46 所示，回归得到的直线与数据点的趋势一致。这说明用最小二乘法所计算出来的参数 a、b 较好地反映了数据的分布，这是一条最佳的拟合线。

同理，我们也可以利用最小二乘法来计算车流量的 a、b 值，进而预测图 2-45 中某路口 8 点至 9 点的车流量。

小贴士

当 J（a，b）最小时，可得

$$a=\frac{m\sum x^{(i)}y^{(i)}-\sum x^{(i)}\sum y^{(i)}}{m\sum (x^{(i)})^2-\left(\sum x^{(i)}\right)^2}$$

$$b=\frac{\sum (x^{(i)})^2\sum y^{(i)}-\sum x^{(i)}\sum x^{(i)}y^{(i)}}{m\sum (x^{(i)})^2-\left(\sum x^{(i)}\right)^2}$$

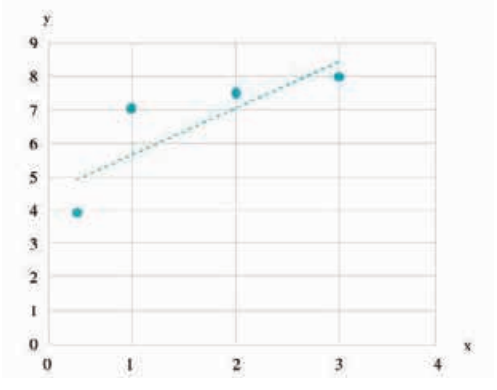


图 2-46　例题中线性回归的结果

活　动

5.2 参考图 2-47，尝试利用线性回归的方法对图 2-45 上的某路口车流量进行预测，计算出回归直线的参数 a、b 的值，并算出 8 点到 9 点预计的车流量（实际值为 1890）。对比一下预测值和实际值的差别，说说你对这个差别的思考。

5.3 尝试了解 Python 中的 Pandas、sklearn 等工具包。利用它们，用计算机模拟的方式重新计算 5.2 的问题。运行配套资源中的代码最终得到预测值和回归后的图形。

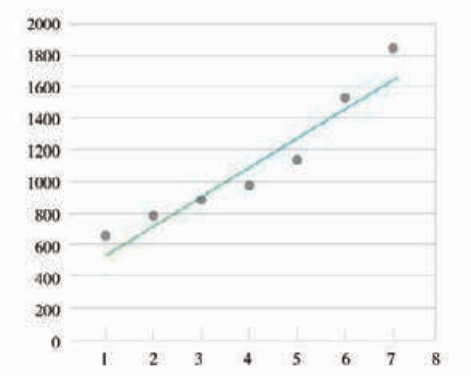


图 2-47　某路口车流量线性回归的结果

3. 根据路况进行智能决策

智能车的智能决策模块可以根据目标以及问题特征进行决策。进行路径规划时，智能决策模块需要对有关路况的各个特征进行判断，这些特征包括施工情况、所处区域的位置情况、通过的时间点等，如图 2-48 所示。只有综合考虑所有相关特征才能准确地进行决策。

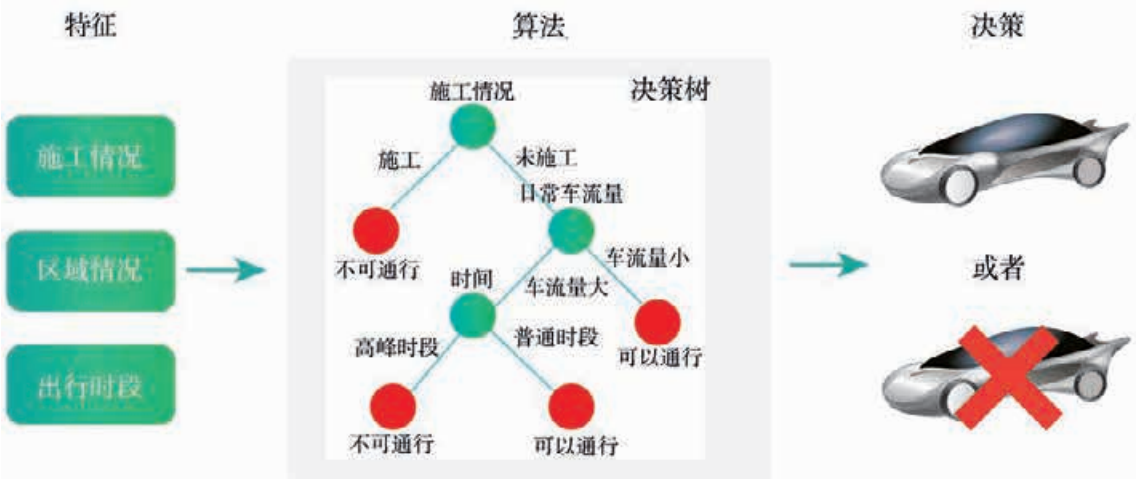


图 2-48　典型智能决策系统模块图

（1）了解决策树算法

为了综合考虑多种特征进行决策，最简单的做法是对所有特征逐个进行判断。决策树（decision tree）算法正是遵循这样的思路。

决策树算法从“树根”开始利用某一特征将样本分类，接着根据分类结果“生长”出数条“树枝”，再利用其他特征进行分类。这样反复“生长”，直到长出“叶子”，即最终的分类结果。决策树用图形来表示即是由多个判断节点组成的（倒置的）树形。

以图 2-49 的决策树为例，它的基本原理是：首先选取施工情况作为判断变量（根节点，root node），根据是否施工，决策树生成两个分叉，也就是树枝。由于只要道路施工，无论其他因素如何，道路均难以通行，因此走到“施工”这个分叉的所有样本，都会直接被判为“不可通行”。如果道路未施工，则无法直接确定可不可以通行，这时需要利用另一个特征属性进一步判断。比如考虑道路日常车流量，如果该区域车流量小，那么不需要考虑其他条件或者特征属性，均可以输出决策“可以通行”。若该区域车流量大，需要接着判断其他特征属性。样本被最终判定为“不可通行”或“可以通行”的子节点（child node）无法再继续生成分叉，被称为叶节点（leaf node）。我们把可以继续形成分叉（即拥有子节点）的树节点称为“非叶节点（nonleaf node）”。

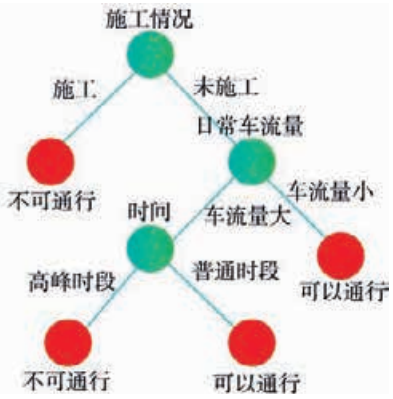


图 2-49　决策树示例

当输入一个样本 $x=(x_1,x_2,\cdots,x_D)$ 后，决策树算法会根据每个节点预设的特征属性 x_i 以及分叉生长规则走到下面一层的节点，接着根据这个新寻到的节点的预设特征属性以及分叉规则，继续往决策树的下一层走，直到某个叶节点，整个行程结束。

每个样本依据其特定的特征取值，在决策树上都有唯一确定的路径（自根节点至叶节点）。在叶节点上，完成最终的决策，如决定是否可通行。

（2）了解决策树的构建方法

决策树需要根据特征属性设置节点，每个节点提一个问题，通过判断将数据分为两类，直到不可再分为止。目前，通常采用下面两种方法来确定特征属性，构建决策树。

① 使用专家规则

在一些情况下，如果问题相关的特征类型不多，而且我们又具备一定的专家知识，那么可以通过这些专家知识，“手动”地建立一棵决策树。

具体实施中，可以人为地规定在生成决策树时所考虑特征的先后顺序，同时对每个分叉节点人为地设置我们认为最合适的分叉判别条件。例如在上述的例子中，一般认为只要

道路施工就应设为不可通行，因此应该先判断道路施工情况，然后考虑其他因素。使用专家规则构建决策树时，所有特征属性的先后判断顺序，以及每个节点分叉条件，都需要人工事先规定好。

② 使用数据训练

在很多情况下，我们并不具备专家的能力。例如某个分类问题有 1000 多种特征属性，这时，人工设置分叉条件显然很难做到。在这种情况下，可以让机器使用大量历史数据（带有最终分类标签的样本）来自动构建决策树。这种方式通常称作决策树训练。

决策树训练通常是逐层地通过选择某个属性，设置某个阈值或分叉条件来训练每个节点。

图 2-50 展示一个通过数据训练生成决策树的例子。表 2-4 中包含了判断某道路是否可通行的历史记录，共有 5 例可以通行和 7 例不可通行的样本。可以发现，凡是包含“施工”这一特征的样本均不可通行，因此先利用施工情况进行判断。这时“施工”分支成为了叶节点，而“未施工”分支还有 8 例样本。再利用预计车流量，将“未施工”分支中可以通行和不可通行的样本分开。当预计车流量小于 79 时，样本均为可以通行，否则均为不可通行，因此将阈值设置为 79。至此所有数据均已到达叶节点，决策树生成完毕。

小贴士

决策树训练的具体过程如下：

① 初始化一个根节点。

② 寻找一个合适的特征属性作为该节点往下分叉的判别变量，并且对该变量设置一个分叉判别函数。

③ 根据判别函数对数据进行分析，将当前节点上的数据往下面的子节点推送，并重复②的操作。

④ 当数据达到某个节点无法继续分类（全属于一个分类标签），则这个节点成为叶节点，不再继续生成子节点。

表 2-4　判断某道路是否可通行的历史记录

特征属性		决策结果
施工情况	预期车流量	是否可通行
施工	48	不可通行
未施工	80	不可通行
施工	83	不可通行
施工	29	不可通行
未施工	71	可以通行
未施工	78	可以通行
未施工	47	可以通行
未施工	87	不可通行
未施工	68	可以通行
施工	75	不可通行
未施工	73	可以通行
未施工	90	不可通行

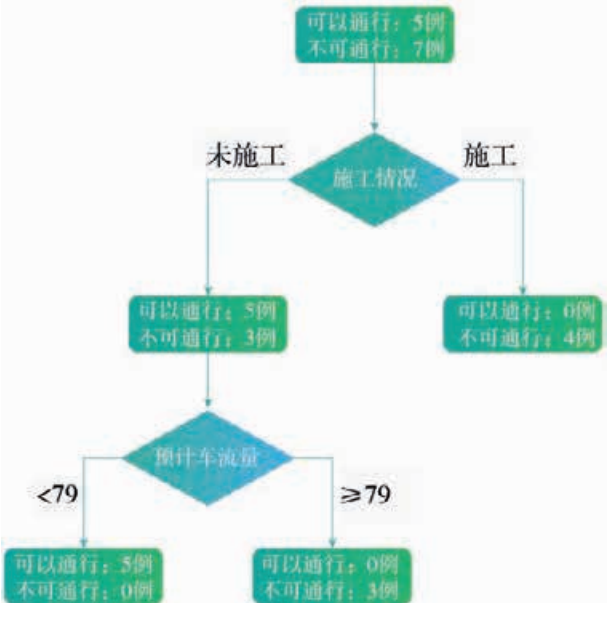


图 2-50　利用表 2-4 的历史数据生成决策树

4. 搜索最佳路线

利用决策树算法对道路进行筛选之后，需要在剩下的“可以通行”的道路中对从起点到终点的路径进行搜索，选择其中最优（最短）的一条。这时，路径搜索系统就发挥作用了。

（1）了解基本的搜索算法思想

为了便于搜索，首先将地图简化，用方格表示地图中的区域，每个方格的状态只有可以通行和不可通行两种。存在障碍物和智能决策排除的道路均标为不可通行。每一步可以从一个方格移动到相邻的可以通行方格。我们将方格称为图中的节点。智能车所在的起点与终点公园之间隔着一条河（图2-51），车子必须从桥上通过，这种情况可以抽象成图2-52的样子（深灰色格子为不可通行，其他格子为可以通行）。

图2-51　一张简单的地图

图2-52　抽象后的地图

路径搜索算法在地图上进行搜索的过程类似于走迷宫。要找到最短的路径需要策略。一般有两类搜索算法：一是盲目型搜索，二是启发式搜索。

盲目型搜索分为两种：深度优先搜索（depth-first search）和广度优先搜索（breadth-first Search）。

深度优先搜索（图2-53）算法的思路是一条路走到底，尽可能地往纵深方向的节点走，直到终点为止，然后再回头尝试其他走法。经过多次尝试之后能够找到一条比较短的路线，但是为了保证找到最短路线，必须走遍所有的路线。

广度优先搜索（图2-54）算法的思路是先尽可能横向探索周边的区域，若没有找到终点，则到下一层的节点继续横向探索周边的区域，直到探索到终点为止。这样可能要探索很多地方，但不一定需要走遍地图就可以找到最短的路线。

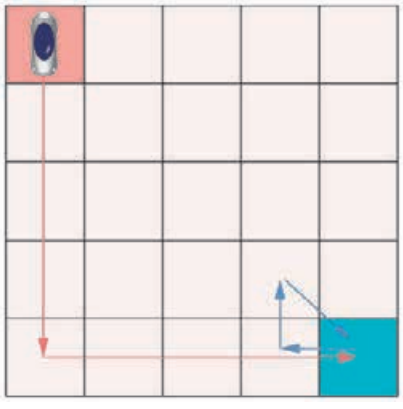


图2-53　深度优先搜索

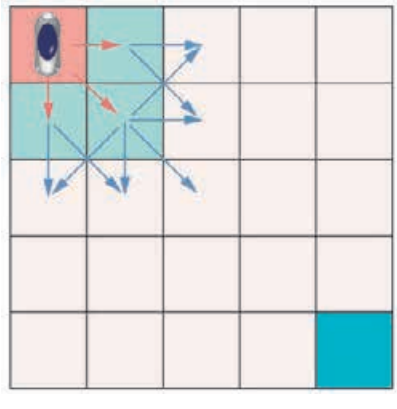


图2-54　广度优先搜索

小贴士

启发式搜索也称为有信息搜索，它利用了问题中的额外信息来指导搜索方向。尽管启发式搜索有时也会出错，导致沿着错误的路径进行探索，但一般而言启发式搜索仍然能够达到远超盲目型搜索的效率。

启发式搜索（heuristic search）算法与盲目型搜索不同，它像明眼人走迷宫，有信息提示它终点在哪，它会优先朝着终点的方向移动或探索。由于有外部信息的支持，所以每一步可以选择较好的方向，最终用较少的步数找到最短路径。一般启发式搜索的效率大大高于其他搜索方法。

（2）了解一种启发式搜索——A* 算法

A* 算法是一种启发式搜索算法。它像广度优先搜索那样先对周边的区域进行探索，但在探索的时候不是每次都向所有可能的方向探索，而是优先朝着最有可能位于最短路径上的位置探索。该算法能够跳过一些明显较差的区域，缩小搜索范围，进而提高搜索效率。它的流程如图 2–55 所示。

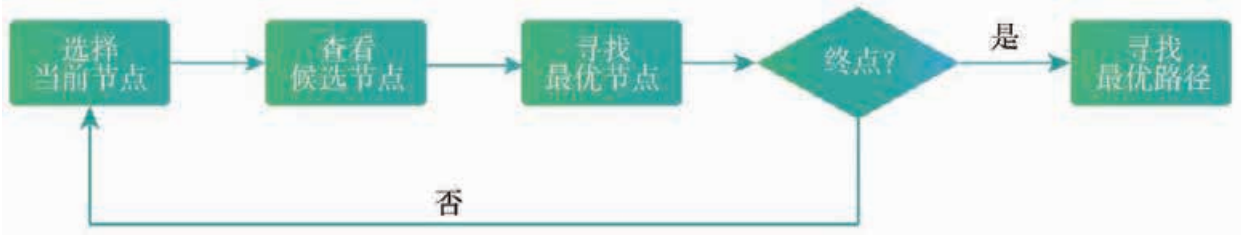


图 2–55　A* 算法流程

A* 算法是为每一步选取与已探索区域相邻节点中的最优节点，并探索该最优节点。障碍物和已探索过的节点不再探索。重复以上过程，直到探索到目标节点，找到最优路径。

为了寻找到最优的下一步（最优节点），需要估计经过某候选节点的从起点到终点的路径总长度。这一路径可以分为两部分，即从起点到该节点的路径和从该节点到终点的路径。用公式表示为：F=G+H。其中 F 是路径总长度，G 为起点到该节点的最短路径长度（已知），H 为该候选节点到终点的路径长度（估计）。与探索区域相邻的区域中 F 值最小的节点，即为我们下一步探索的最优节点。一般情况下，距离终点越近的节点到终点的路径长度也越短，为了便于估计，可以令 H 为该节点到终点的直线距离。如图 2–56 所示，左上角粉红色节点为起点，右下角蓝色节点为终点，那么绿色实线的长度为橙色节点的 G 值，蓝色虚线的长度为橙色节点的 H 值。可见，越接近图中对角线的橙色节点，蓝线和绿线的总长度也越短。因此，接近对角线的节点比起其他节点更有可能位于最佳路径上。

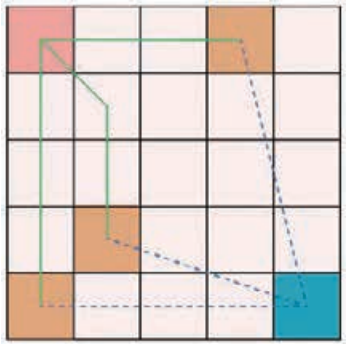


图 2–56　G 值和 H 值的示意图

（3）使用 A* 算法搜索最短路径

在图 2–57 所示的抽象地图上，红色表示起点，深蓝色表示终点，浅蓝色表示候选节点，橙色表示当前节点，深灰色表示障碍物和已探索节点。图中数字为该节点的 F 值。

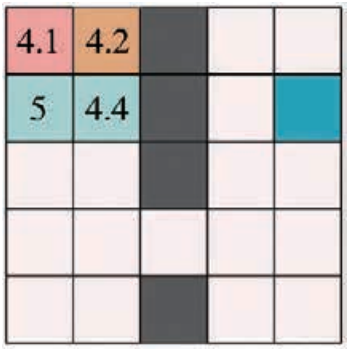


图 2–57　使用 A* 算法搜索最短路径之 1

参见 P64 知识链接“A* 算法的路程估计”

首先，将起点作为第一个已探索节点，把与它相邻的节点加入候选列表，并计算它们的 F 值。

起点右侧的节点具有最小的 F 值，所以第二步选择该节点为当前节点（橙色节点）。将与橙色节点相邻的节点加入候选列表。这一步没有新的节点加入候选列表，因此下一步选择 F 值仅次于当前节点的节点，即起点右下方的节点。

第三步，将起点右下方节点设为当前节点，把与该当前节点相邻的节点加入候选列表，并计算 F 值，如图 2–58 所示。

第四步，选择 F 值最小的起点下方的那个节点（如图 2–59 中的橙色节点），并更新候选列表。

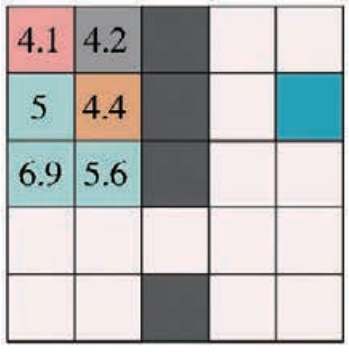


图 2–58　使用 A* 算法搜索最短路径之 2

图 2–59　使用 A* 算法搜索最短路径之 3

反复以上的探索过程，直到探索到终点为止，即可找到从起点到终点的最短路径。

本项目中，为了实现路径规划，我们先采用线性回归预

小贴士

图中节点的 F 值计算方法：水平和竖直方向的一步距离为 1，对角线方向的一步距离按 1.4 计算。

第四步中并没有新的节点被加入候选列表，但是起点下方第二个节点的 F 值改变了。这是因为原本的 G 值是按照起点—右下—左下的路线计算的，而探索了起点下方节点后就可以选择直接向下的路径，G 值可以得到减小，因此该节点的 G 值和 F 值被更新。

测道路的车流量，再通过决策树综合多种因素，排除难以通行的道路，最后利用启发式搜索找到前往终点的最短路径。这三种算法在社会其他领域也有着非常广泛的应用(图 2–60)。

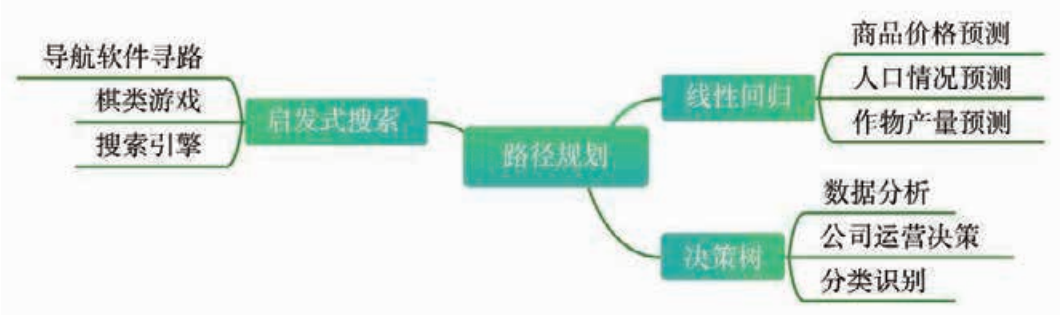


图 2–60　路径规划所涉算法的社会应用

本项目中，我们只是了解了单变量线性回归的应用。但在实际生活中，影响车流量的因素还有很多。社会生活中的问题往往更加复杂，可能需要用多变量线性回归来进行预测。线性回归能够反映数据的变化趋势，经常用于预测各种指标，例如商品价格变动趋势、人口情况发展趋势等。

决策树算法能够综合多种因素作出决策，常常用于分析大量数据，以辅助公司运营决策。另外，决策树本质上是对样本进行分类，因此它也可以用于一般意义上的分类识别，例如图像和文本的分类等。

启发式搜索能够高效地从数量巨大的可能情况中找出最优解，这使它在智能路径规划中起到不可代替的作用。除此之外，像需要在非常多的走法中找出最优解的棋类游戏和需要在网络中高效地爬取数据的搜索引擎等应用领域中，启发式搜索也发挥了作用。

活　动

5.4 尝试使用 A* 算法，在本项目的地图（图 2–52）上找出从红色节点到蓝色节点的最短路径。算出各个必要节点的 F 值，并最终找出一条最短的路径。

完成后与图 2–61 对照验证其正确性，并思考：本例中最短路径不止一条，为什么最短路径选择的是图中的这一条？

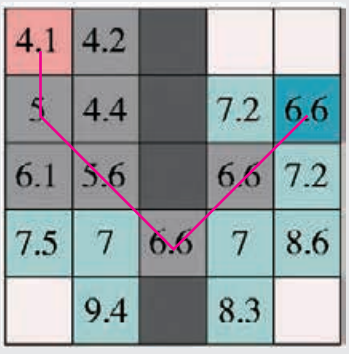


图 2-61　最短路径结果

5.5 尝试用计算机运行决策树算法和 A* 算法。配套资源中给出了运行程序所需的函数、保存地图信息的“map.txt”和所需代码。指定决策策略后，体验用决策筛选和路径搜索找出从起点到终点的最短路径的过程。

知识链接

多变量线性回归及梯度下降

一个路口的车流量不仅仅与时间有关，还与附近居住的人口、城市的繁华程度等因素有关。将两个或两个以上的影响因素作为自变量，并推测它们与因变量变化之间关系的回归，称为多变量回归。如果它们之间的关系是线性的，则称为多变量线性回归。具体地说，就是要找到如图 2-62 所示的一个平面，使所有的真实数据点到这个平面的距离之和 J 最小。

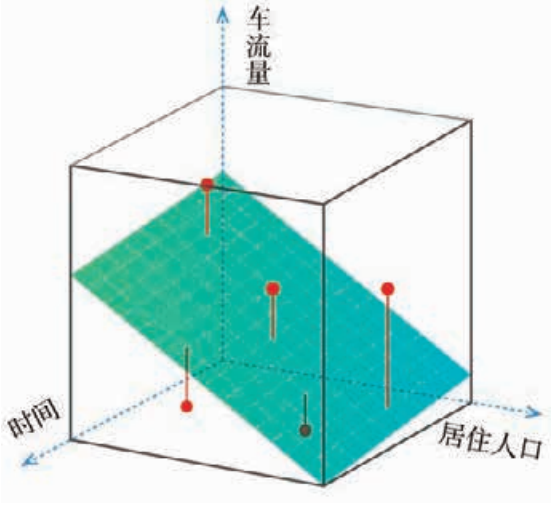


图 2-62　多变量线性回归

人们常用梯度下降法来帮助找到这个平面。梯度下降法是沿梯度相反方向（下降最快方向）求解函数的极小值。梯度下降的算法有两个可调参数，学习率 η 和迭代次数 N。学习率 η 决定了每次梯度下降的幅度，迭代次数 N 决定了迭代的次数。如图 2-63 所示，假设从 A 点出发，要找到一条快速到达最低值 D 点的路。梯度下降的思想是环顾四周找到一个下降最快的方向然后走一步。接着，再环顾四周沿着新的下降最快的方向走一步。不断这样重复，直到发现周围都比我们高了，那么就认为走到最低点了。当然有时候，如果学习率和迭代次数设置不当，如从 B 点出发，也是按照当前下降最快的方向下降，却可能落到局部最低点 C。正常情况下，可通过多试几组学习率和迭代次数来尽量避免落入局部最低点。最后，通过设置好梯度下降的学习率和迭代次数，不断重复下降的过程，就能找到最优的参数使距离之和 J 最小。

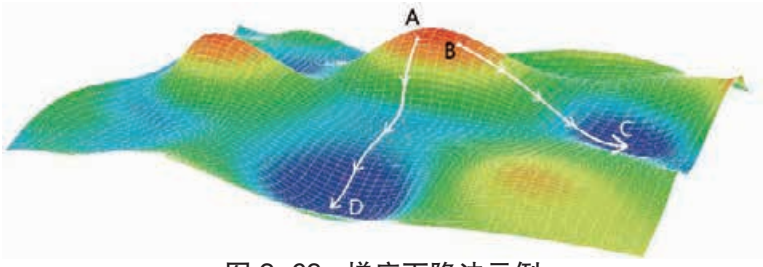


图 2-63　梯度下降法示例

A* 算法的路程估计

在 A* 算法中，我们必须对候选节点到终点的路径长度进行估计。由于无法得知具体路线，通常都用距离来估计路程。项目正文采用的是直线距离，又称为欧氏距离（Euclidean distance），但事实上还有两种距离在路径规划中应用得更为广泛：曼哈顿距离（manhattan distance）和对角线距离（diagonal distance），如图 2-64 所示。

曼哈顿距离是指两点在水平方向和竖直方向的距离之和，也就是只能进行水平和竖直两个方向移动时的最短路径长度。由于城市中的道路通常横平竖直，曼哈顿距离用于估计城市中的路径长度较为准确，因此也被称为城市街区距离。

和曼哈顿距离相比，对角线距离还允许沿方格对角线方向的移动，这和正文例子中每一步的移动方式一致，因此是最适合本项目例子中路程估计的距离。下图给出了欧氏距离、曼哈顿距离和对角线距离下的最短路径。

欧氏距离　　曼哈顿距离　　对角线距离

图 2-64　三种距离下的最短路径

拓展阅读

自回归滑动平均 (Auto Regressive Moving Average，ARMA) 模型

对于如图 2–65 所示的复杂时间序列分析和预测，通常需要引入自回归滑动平均之类的更复杂模型来进行预测。

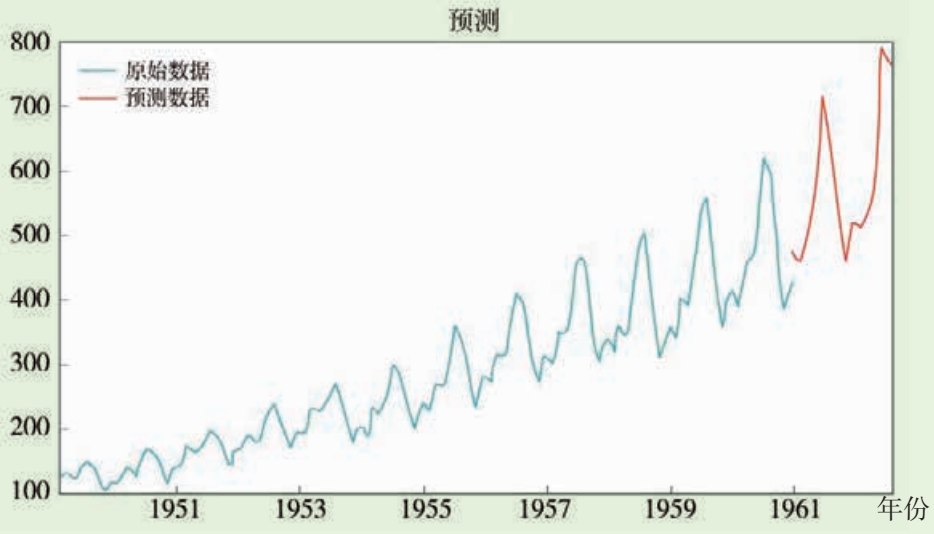


图 2–65　某地区火车乘客数量统计图

我们可以直观地看到这个地区火车乘客数量变化有两个特点：

（1）趋势：从长期来看，乘坐火车的人数是不断增长的。

（2）周期性：在春节等假期期间，乘坐火车的人数会增加。

对于具有这样特性的时间序列，自回归滑动平均模型处理方法是先将趋势和季节性等影响序列的因素通过建模的方法移除，形成一个简单的时间序列，然后对这个简单的时间序列进行预测，将得到的预测值再加上建模后的季节性和趋势的约束，得到最终的预测值（图 2–65 中红色部分）。从图 2–65 中可以看到，预测值很好地延续了历史数据的周期性和趋势。

自回归滑动平均模型广泛应用于经济活动和自然现象的预测。比如，用于对国际原油价格的预测、对具有季节性变动特征的商品销售量的预测、对小麦产量的预测等。

——摘译自 Aarshay Jain.A comprehensive beginner´s guide to Create a Time Series Forecast(with codes in Python).2016

项目六

让智能车识别道路障碍物

——认识人工神经网络与深度学习

要让智能车能够上路，并且在遇到行人时，可以自主地减速停车，待行人通过后再启动，就必须让智能车能识别障碍物，其本质是对障碍物的图像进行识别，如图 2-66 所示。因此，智能车需要具备物体识别系统。目前大多数接近实用的人工智能物体识别系统，如车牌识别、人脸识别、交通标志识别等，都基于人工神经网络和深度学习。

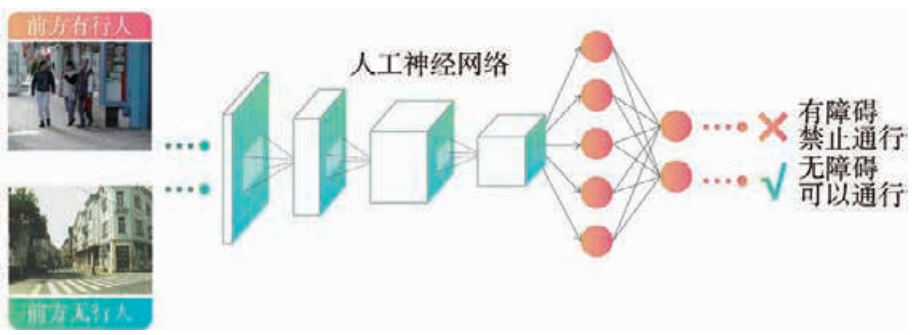


图 2-66　基于人工神经网络的智能车障碍物识别过程

项目学习目标

在本项目中，我们将通过探究如何让智能车识别障碍物，来学习人工神经网络和深度学习的思想、方法及相关技术。

完成本项目学习，须能回答以下问题：

1. 什么是人工神经网络？其基本原理是什么？
2. 什么是深度学习？其基本原理是什么？
3. 人工神经网络如何进行图像识别？
4. 深度学习和人工神经网络之间的关系是什么？

项目学习指引

智能车上路时，必须采集车辆周边的实时信息，对周围的每个物体、交通标识进行辨认、分析。如果车辆前方有障碍物，首先要在车辆运动的前提下，确定障碍物是人还是物，是静止的还是运动的。这些都需要能在极短的时间里加以识别、判断。由于天气、道路环境等条件复杂，这种识别和判断还会受到光线变化、遮挡等问题的干扰。

相比简单的静态人脸识别，运动状态下的智能车对周围障碍物的识别难度高得多。在过去数十年里，一些研究人员致力于从人脑的神经系统汲取灵感，设计人工神经网络来解决复杂的图像识别问题。目前，人工神经网络和深度学习技术在模拟人类认知的道路上得到深入发展，也被广泛用于解决智能车障碍物的识别问题上。

1. 初识人工神经网络

（1）人工神经网络的结构

人工神经网络在结构和原理方面模仿了人类的大脑神经网络。图 2-67 展示了现代主流人工神经网络的基本结构。图中每个圆圈代表一个神经元，纵向的几个神经元组成一层神经网络，各层神经网络间互相连接（可以顺次，也可以不顺次），上一层的输出作为下一层的输入，这样互相联系，组成最终的神经网络。

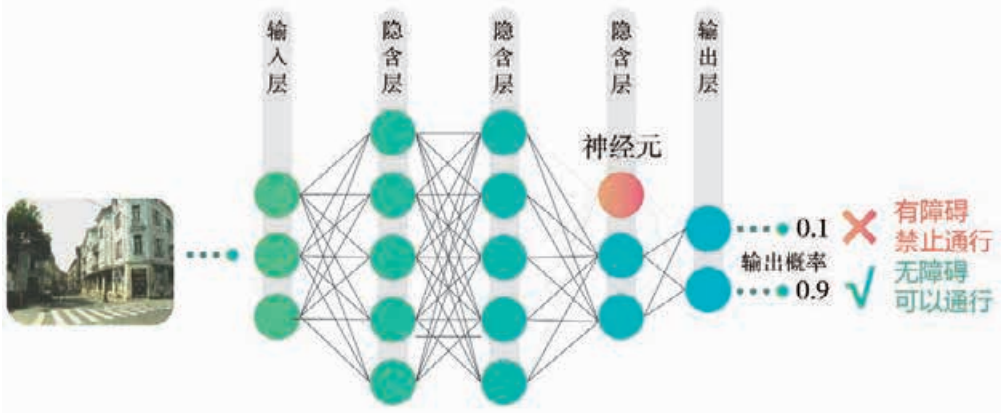


图 2-67　人工神经网络组成结构示意图

一个完整的人工神经网络通常包括一个输入层、若干个隐含层（也称为中间层）以及一个输出层。输入层的神经元与输入信息的所有维度相连接，例如图片识别中，每个输入层的神经元均与每个像素点连接。输入层与隐含层的所有节点两两相互连接。输出层的节点通常指示所要识别的语义标签。对于图片识别来说，如果有 C 个类别（图 2–67 中 C=2），输出层通常就有 C 个神经元。每个神经元输出的值在 0 与 1 之间，表明所输出的图片属于某个类别的概率。这种上一层所有节点都和下一层任意节点相连的人工神经网络也称作全连接神经网络。以行人识别为例，其本质就是图像识别：神经网络判断当前路段是否有行人，实际上就是用许多神经元处理输入的当前道路图像的像素，并从中提取特征信息，进而作出判断。

小贴士

分层是神经网络的一个重要概念。神经网络的层级结构具有非常强大的特征抽象能力，是决定神经网络性能优越程度的关键。

参见 P77 知识链接“行人检测”

（2）人工神经网络的基本单元：神经元和激活函数

人工神经网络的提出最初是受到人脑信息处理方式的启发（图 2–68）。人脑神经网络的基本单元（神经元）负责信息的感受、激活和传输。人工神经网络模仿人脑的神经元构建了相似的基本单元，用于承担信息的感受、激活和传输三大基本任务。

小贴士

人工神经元只是人脑神经元的数学模型。在实际使用中，主流的人工神经网络与人脑的神经网络并没有直接关系。

图 2–68　人脑神经元与人工神经网络神经元对比

人工神经网络中的神经元将输入信息线性组合后（感受），经过非线性变换（激活），再传给下一层（传输）。如将输入写作（a_1，a_2，…，a_N），将输出写作 a_{out}，那么他们之间的关系可以用数学表达式表示为：

$$a_{out}=g\left(b+w_1a_1+\cdots+w_Na_N\right)$$

在这个数学表达式中，w_i 和 b 是神经元的参数，代表了信息的处理过程；函数 g 是非线性函数，也称激活函数。

小贴士

激活函数（activation function）在提升人工神经网络性能方面起着至关重要的作用。激活函数可以认为是人工神经网络必不可少的一部分。这里“激活”的意思实际上是指对于不同输入范围的信号，产生不同的响应机制。

下页图 2–69 为四种常用的激活函数。ReLU 函数对于负向信号没有响应，而完全保留了正向信号；Sigmoid 函数则不同，它对于幅度较小的信号有响应（在 –0.5 与 0.5 之间近

似线性），而对于幅度过大的信号则相应饱和（输出近似为 1）。

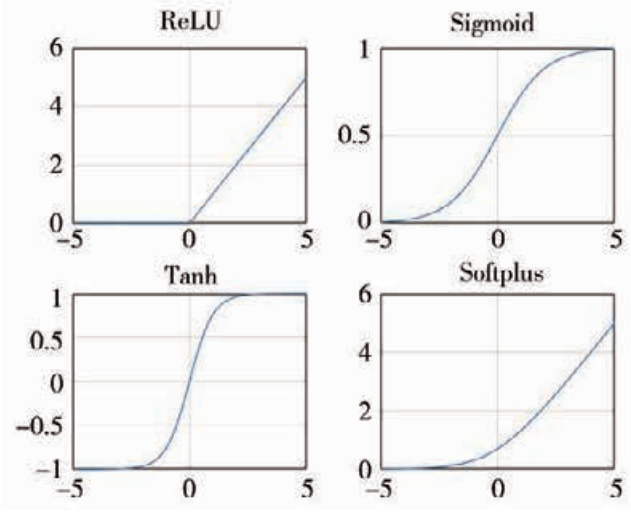


图 2-69　四种常用的激活函数

（3）人工神经网络的功能：前向计算

如果输入一张有障碍物的道路图片，人工神经网络如何输出“前方有障碍物”的标签呢？这可以用到人工神经网络的主要计算功能——前向计算。人工神经网络计算的发生顺序是不断地将刺激由一层传向下一层，即按照从输入层，经过隐含层，最后到输出层这样向前的顺序。而每个节点的输出值计算过程就是神经元的处理过程。通过这种方法，一层层不断地运算，最终得到输出层结果。

人工神经网络的前向计算可以看作是一种静态非线性映射。通过简单非线性处理单元的复合映射，可获得复杂的非线性处理能力。实际上，如给予人工神经网络足够多的神经元和层数，人工神经网络能够以任意精度逼近（模仿）任意连续函数，得出足够精确的结果。

（4）人工神经网络的训练方法：反向传播

人工神经网络需要具备最佳参数来进行有效的图像识别，这个寻找最佳参数的过程就是训练。反向传播算法是一种很好的训练人工神经网络的算法。

反向传播以误差减小为准则，旨在得到最优的参数矩阵，进而将多层人工神经网络应用到分类或者回归任务中去。以图 2-70 所示的一个两层人工神经网络为例，该人工神经网络包含两个输入节点 a 和 b，两个隐含节点 c 和 d，一个输出节点 e。相应每个节点的输出是 y_a、y_b、y_c、y_d 和 y_e。相邻两层节点由权重 w 连接，如节点 a 和节点 c 之间的权重用

小贴士

反向传播算法于 20 世纪 80 年代由机器学习研究者们提出。

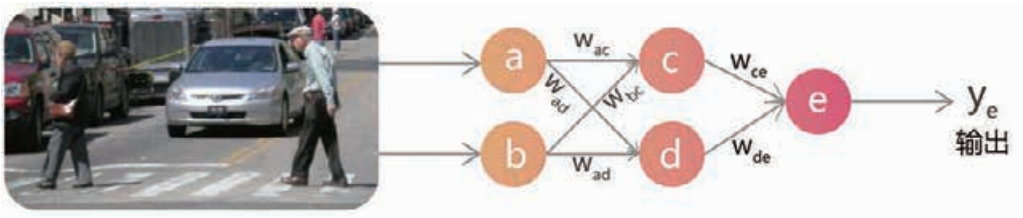


图 2-70　两层人工神经网络示例

w_{ac} 表示。图中实线灰色箭头代表了信息传播方向，即从人工神经网络第一层向前传播至第三层；而反向传播就是按相反方向，即从第三层向第一层不断更新权重参数。

2. 了解深度学习及其基本操作

下面以卷积神经网络（CNN）为例，理解深度学习的基本思想。

（1）深度学习

使用人工神经网络识别车前方障碍物时，我们会发现，一个 640×480 像素的红绿蓝三色图像，包含的像素非常多，这说明其参数量巨大。由于相邻两层节点两两相连，假设第一个隐藏层节点数为 1000000（通常隐含层的神经元数目会更多），仅第一层参数量就达到了（640×480×3）×1000000=9.216×10^{11}。如果算上所有层的参数，整个人工神经网络会非常巨大，给计算和存储带来极大的挑战。随着大数据时代的到来，数据的量和复杂度已超出浅层人工神经网络的学习能力，导致浅层人工神经网络在大型数据集上的表现不能令人满意。这就要求人工神经网络必须越来越“深”。

人类大脑的神经元“各司其职”，即每个区域的神经元只对特定的输入进行响应。同时，人类的神经网络层级连接非常“深”，且随着层级加深，神经元的感受野逐渐变大。浅层的神经元可能只对形状和方向有响应，深层的神经元对方向和运动均有响应，而更深层次的神经元则对复杂的复合运动有响应。通俗地说，就是前一层每个神经元只处理 1cm×1cm 大小的区域，后一层的神经元可处理 10cm×10cm 的区域，感受野不断变大，最后的几个少量神经元可以输出整个图像的抽象表达。换句话说，前面的层级处理比较局部的表观信息，后面的层级处理更加高级的语义信息，最后的层级则完全理解图像的全局语义信息。受人类大脑的神经元深层连接结构启发，人工智能科学家为人工神经网络设计了

小贴士

更新权重参数的目的是将人工神经网络的预测输出与真实值进行比较（如做差运算），然后将误差通过反向传播一层层地向人工神经网络的前面层传播。通过传播的信号，修改网络权重参数，使误差变小。

小贴士

卷积神经网络是深度学习技术中极具代表性的网络结构之一，它在图像处理领域取得了很大的成功，在自然语言处理等领域中也有重要应用。

小贴士

深度学习的“深”既表示人工神经网络的层次多，也表示模型参数多。

小贴士

在图像识别中，感受野（receptive field）用来表示神经网络内部的不同神经元对原图像的感受范围大小。人工神经网络中，指每一层输出的像素点在原始图像上映射的区域大小。

卷积操作和池化操作两大技术，使得深度神经网络变为可能。

（2）深度学习的基本操作——卷积操作

在深度神经网络中，一类很重要的网络结构称为卷积核。卷积核是这样来处理信息的：

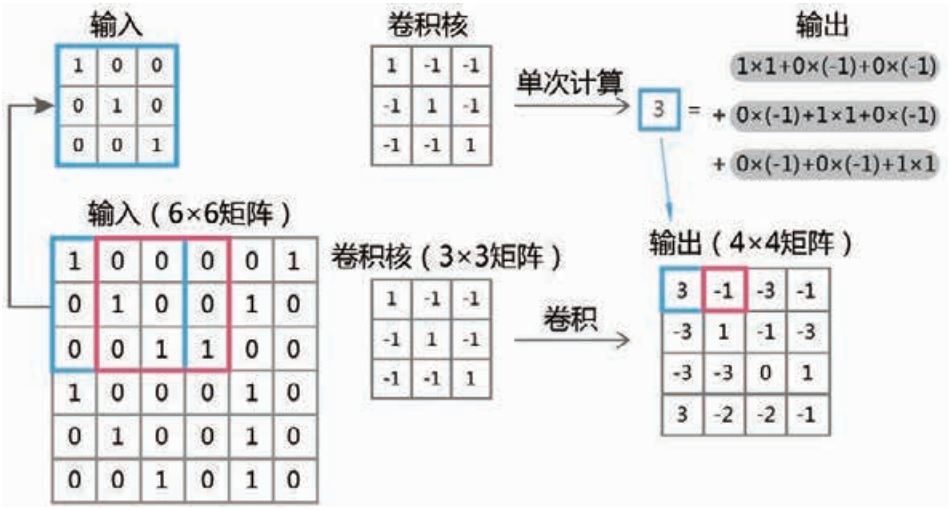


图 2-71　卷积操作图示

卷积核通常是远小于输入图像尺寸的数据矩阵，如图 2-71 所示的卷积核是一个 3×3 的数据矩阵，它与输入图像中的蓝色框内数据进行卷积（卷积核与相应输入位置相乘再相加）操作，得到输出结果 3。然后把卷积核平移到下一个位置（红色框），进行相同的操作，得到下一个计算结果（-1）。以此类推，得到最终的输出结果（4×4 的矩阵）。通俗地讲，卷积核就像一个滑动窗口，沿着表示输入图像的矩阵逐格滑动，进行点乘，最终根据滑动的位置生成相应的输出，这样卷积核的权值被整个图像共享。相对于全连接神经网络，经卷积操作的神经网络的参数量一下子从（6×6）×（4×4）= 576 降低到（3×3）= 9（即卷积核的大小）。像这样包含卷积计算且具有深度结构的神经网络，称为卷积神经网络。

在道路障碍物识别时，卷积核是一张典型的障碍（如图 2-72 中的行人）照片，输入是前方的路况。卷积神经网络就是拿着这张照片（卷积核）在前方道路上不断比对（对应卷积操作时的平移滑动）。如果某一区域内的图像与照片足够相似（输入的某一区域和卷积核足够相似），就判断该区域内有行人（输出一个高响应），否则判断没有行人（输出一个低响应），最终判断出当前道路上是否有障碍（行人）。

数字化学习

上网查找卷积操作的案例，理解其思想。

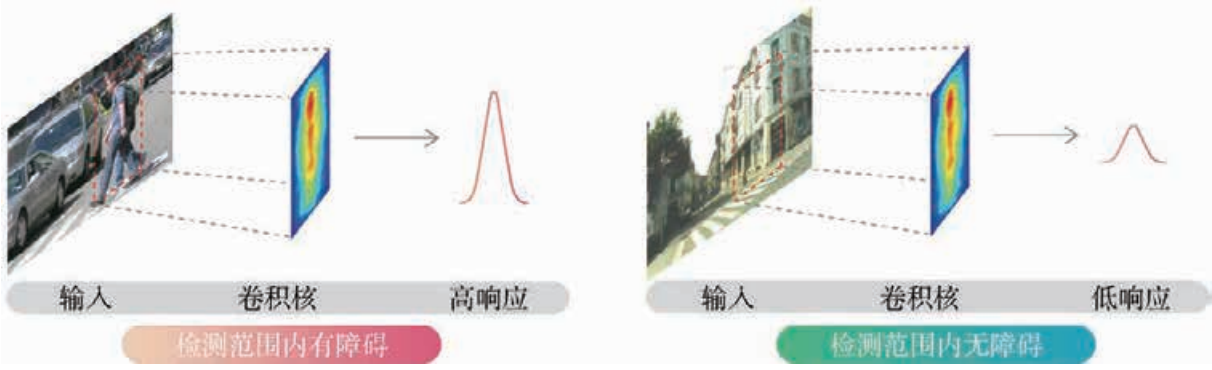


图 2-72　利用卷积操作识别障碍物

（3）深度学习的基本操作——池化操作

卷积操作的引入大大降低了人工神经网络特别是深度神经网络的参数量。在深度学习中，池化操作能使人工神经网络对于信号的处理也跟人脑一样，由低层次局部图像特征的处理，不断向更大感受野、更多语义信息的处理发展。不同于卷积操作，池化操作是一种非线性操作，它同时能够直接改变输出特征的分辨率，将其映射到高维非线性空间。所以池化操作在人工神经网络特征提取中是非常重要的一环。图 2-73 展示了两种常见的池化操作：平均池化和最大池化。

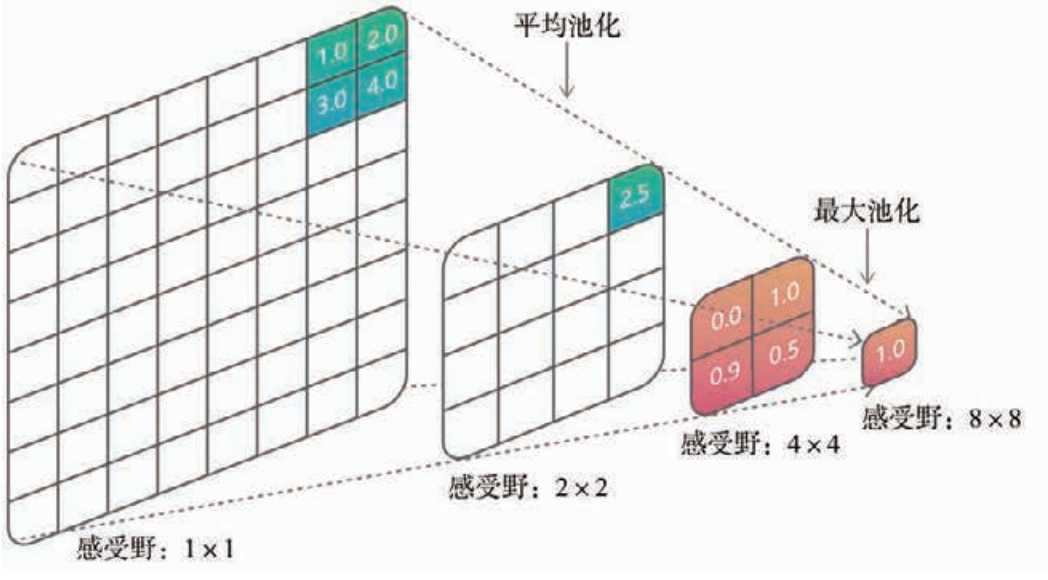


图 2-73　平均池化和最大池化的感受野变化

最大池化是选择特征图中 2×2 区域内最大值作为输出；而平均池化则是将 2×2 区域内的所有值平均后输出。两种池化操作都把输入特征缩小了，同时对其进行非线性映射得到输出，成为神经网络特征选择过程中的重要组成部分。通

过池化操作，神经元的感受野随着层数不断增大。

从池化层的具体操作可以看出，池化操作具有非常直观的物理含义。如图 2–73 所示，在输入特征中每一个色块都可以看作一个“池子”（大小为 2×2），最大池化就是从“池子”中选取最大的那个值作为输出，而平均池化把“池子”中的所有元素值平均后输出。

不同的池化操作实际上是从输入特征中抽取具有不同统计特性的部分，从而强化网络的特征抽取能力。

图 2–74 是一个典型的深度神经网络图示，其中包含了卷积操作和池化操作。从中可以看到，随着层级加深，学习到的卷积核具有越来越明确的语义抽象含义（像素—边缘—部件—轮廓—物体）。现代人工神经网络的成功证明人工神经网络的层级结构具有非常强大的特征抽象能力，而这种抽象能力也是神经网络性能如此优越的关键。

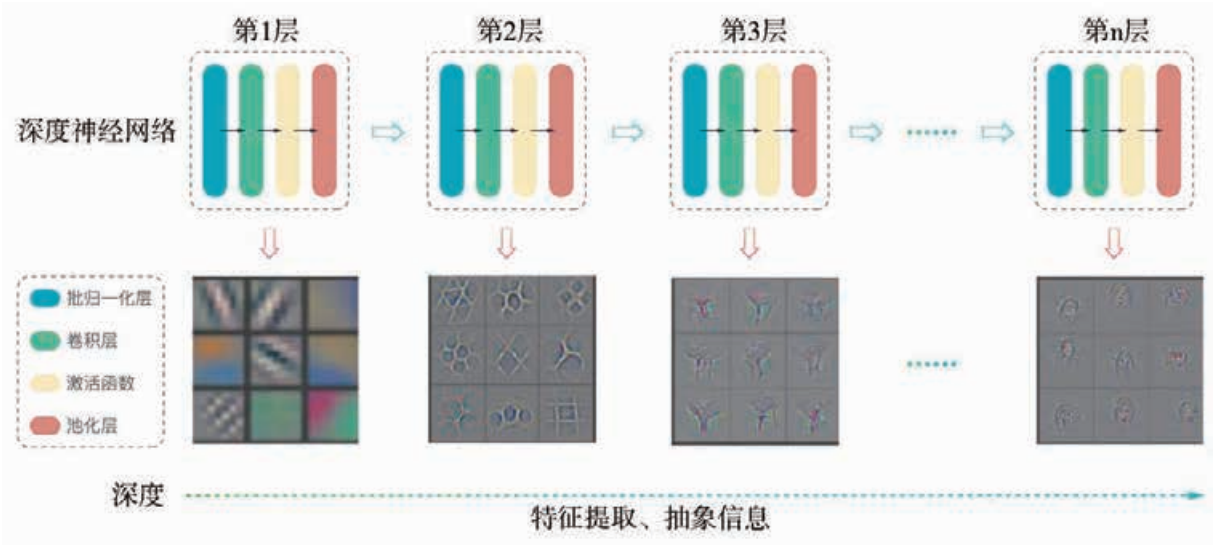


图 2–74　典型深度神经网络图示

3. 探索深度学习的最新发展

（1）深度学习发展中的难题

人工神经网络在 20 世纪就有了较为完备的理论基础，但在其发展过程中遇到很多难题。尤其是随着网络层数的增加，其参数总量也在快速增长，这给人工神经网络的训练带来了困难。

① 过拟合与欠拟合

在深度学习领域中，经常遇到的一个难题就是模型的学习能力与数据的复杂度不匹配，进而导致模型精度下降的现

象。如果模型能力远超数据复杂度，那么在学习过程中模型会将本不是特征的属性（如噪声）当作某种特征，从而遇到测试数据时会因为“过度学习”而导致测试精度下降，我们称之为过拟合（over-fitting，图 2-75）。另外一种情况是模型能力不足以学习到全部特征，某些与精度紧密相关的特征没有被模型学习到，这同样会导致测试的精度下降，我们称之为欠拟合（under-fitting，图 2-77）。如何匹配模型能力与数据复杂度，使其达到最优拟合（图 2-76），避免过拟合与欠拟合，一直是深度学习甚至机器学习领域内的一个重要问题。

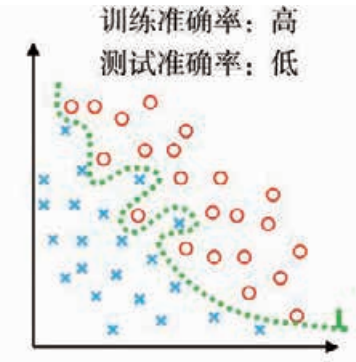


图 2-75　过拟合

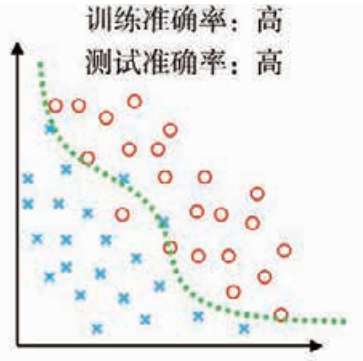


图 2-76　最优拟合

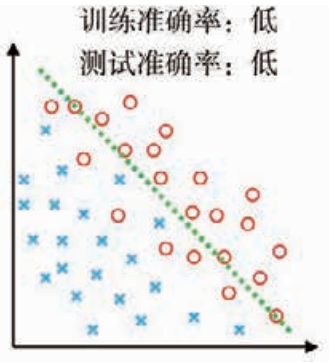


图 2-77　欠拟合

图 2-78　汽车沿山坡下山

图 2-79　汽车驶入平原

② 梯度消失与局部最优解

深度神经网络另外一个需要面对的问题是反向传播算法的失效。反向传播的梯度信号在最初的几层神经网络具有较大的幅度。通常距离输出层最近的几层神经网络可以得到较好的更新。就好像图 2-78 中一辆汽车的目的地在某个地势较低的地方，沿着山坡向下行驶即可，地形能明确地指引方向。

但是由于激活函数往往会限制信号的幅度，导致反向传播的梯度信号也会随着层数的增加而减弱。距离输出层较远的神经网络参数基本得不到更新，进而导致模型精度受损。就好比图 2-79 中汽车驶入平原地形，失去了明确的地形作为指引，行驶方向变得不明确。这就是深度学习中经常遇到的“梯度消失问题”。

类似于梯度消失，深度学习中还会遇到的一个问题是陷入局部最优解（local optima）。这是指模型参数被优化到某个局部范围内的最优值，导致模型的性能无法进一步提高。这好比汽车沿着地形行驶到某一个山谷底部并停了下来，但是它的目的地（全局最优解）并不在这里。如何避免陷入局部最优解也是深度学习需解决的一个关键问题。

这些难题在人工神经网络技术被提出的很长一段时间里都没有很好地解决。但近年来，随着深度神经网络算法和结构得到革新，数据量的迅猛增长，以及以 GPU 为代表的并行计算技术的出现，梯度消失、陷入局部最优解等问题得到一定程度的缓解，从而使得深度学习的理论得以发展。

（2）深度学习的新技术和新机遇

① 卷积神经网络共享权重和 GPU 加速训练

为了解决参数过多的问题，人们首先提出使用卷积神经网络实现参数共享。卷积操作是卷积核在输入信息上进行空间滑动，这其实就是实现了输入信息对于卷积核的共享。相比于全连接神经网络，卷积神经网络大大降低了参数总量。同时，以 GPU 为代表的并行计算技术的出现，实现了网络训练的并行化。有并行计算的帮助，我们可以高效地处理多个网络参数和数据。图 2-80 展示了近几年在 GPU 驱动下深度学习计算速度的提升情况（第一列为 CPU 驱动下的深度学习计算速度，假设为 1，则 2015 年 12 月时，主流 GPU 驱动下深度学习计算速度为 53）。

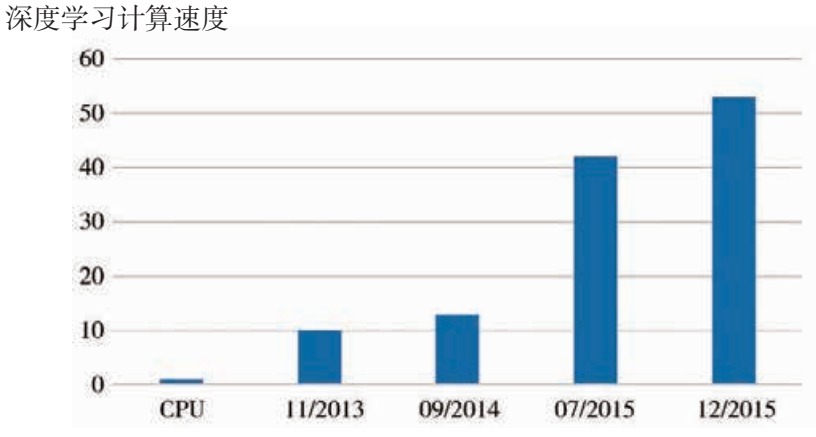


图 2-80　GPU 驱动下深度学习计算速度的提升情况

② 算法和结构上的革新

为了解决梯度消失难题，最近几年人们又提出了很多算

法和结构上的革新。如表 2-5 所示，人们使用了全新的激活函数，同时增加了 Dropout 层，还有批归一化处理技术和首次提出的跨层连接技术。

表 2-5　深度学习相关技术革新情况表

解决的问题	采用的技术	原理简介
过拟合	Dropout	多模型融合
梯度消失	新的激活函数	非饱和函数保留梯度
	跨层连接	多通路保证梯度幅度
	批归一化处理	统一数据分布

人工神经网络技术提出初期，激活函数仅有 Sigmoid 一种。为了解决优化算法的失效难题，人们提出了更为简单却有效的激活函数（如 ReLU、Softplus 等）。它们很好地保证了误差信号在向后传播过程中的影响力，使人工神经网络的性能得到保证。以 ReLU 函数为例，它和 Sigmoid 函数的最大不同在于正向信号（大于 0 区域）不饱和（输出范围不限），这样可以有效保证梯度信号在经过非线性层时的幅度，从而缓解梯度消失的问题。

随着新的数据处理技术的出现，如批归一化处理等，误差在后向传播过程中的有效性得到进一步加强。直观地说，批归一化处理技术保证了数据分布被归一化到一个合理的范围内，在这个范围内可以将误差信号的影响力最大化，使得深度神经网络的参数更新成为可能。

跨层连接是 2015 年提出的。跨层连接实现方式较为简单（不相邻层之间存在信息通路），且经过大量实验证明它在缓解梯度消失方面非常有效，现在已经成为解决此类问题的重要手段。

在新型优化算法、海量数据和并行计算三驾马车的帮助下，以深度神经网络为主的深度学习在近几年大放异彩。如图 2-81 所示，从 2010 年开始，人们不断尝试提高人工神经网络的层数，得到的网络性能不断攀升（错误率不断降低），甚至已经超越人类专家在 ImageNet 数据集上的表现（TOP5 错误率 5.1%）。不仅如此，深度神经网络在物体检测、图像分割、图像生成等领域也有非常成功的应用。

数字化学习

上网了解 ImageNet 和 TOP5 错误率的含义。

参见 P78 知识链接“深度学习最新进展”

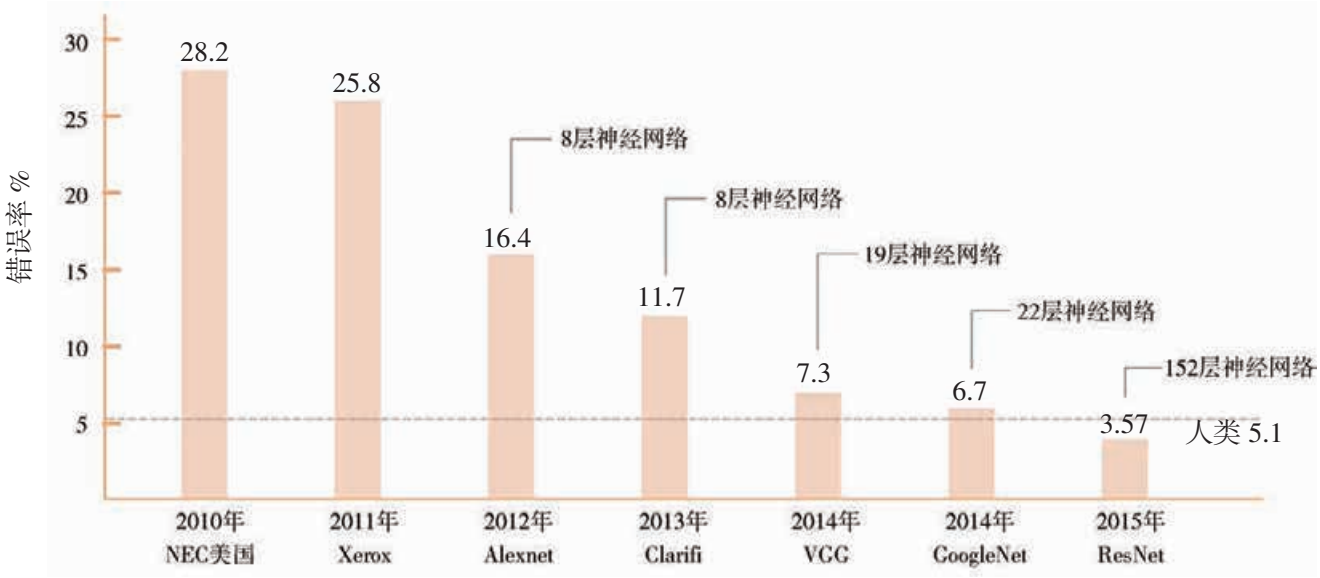


图 2-81　历年 Image Net 竞赛 TOP5 错误率对比

活　动

6.1 结合配套资源，尝试利用深度神经网络检测样例图片中是否有行人。

活动步骤为：（1）在计算机上搭建基本的深度学习运行环境。（2）正确运行所提供的 Python 程序（detector.py），载入模型并读入图片。（3）观察输入某一图片时的相应输出，检验深度神经网络的输出是否正确。

输出结果的格式为：（1）如果判断含有行人，则输出“xxx.png(图片名称) contains person”；（2）如果判断没有行人，则输出“xxx.png(图片名称) does not contain person”。

6.2 在活动 6.1 的基础上，结合配套资源，尝试观察深度神经网络对图片的特征提取结果。

活动步骤：（1）回顾所学知识，尝试分析所抽取的特征具有的特点。（2）运行提供的 Python 程序（feat_vis.py），正确载入模型和读取样例图片。（3）观察相对应的特征提取结果并分析其特点。

注意：提取出的特征被分别保存在 6 个不同的图片中，同时对比观察这些图片会有更加直观的感受。

知识链接

行人检测

图 2-82 展示了将行人检测的特征进行可视化结果。在输出端，蓝色越深表示神经网络响应越小，红色越深表示神经网络响应越大。当有行人存在时，神经网络的不同层对行

人的响应部位不同，也就是说有些层负责检测躯干，有些层负责检测头部，而有些层负责检测四肢等。躯干、头部、四肢等都是人体的典型特征，将这些特征组合起来，在输出端就可以看到一个大致的人体形状。可见，深度学习范畴下的神经网络学习过程是一个抽象和理解目标物体典型特征的过程，可以认为深度学习赋予了人工神经网络很强的自主特征学习能力。

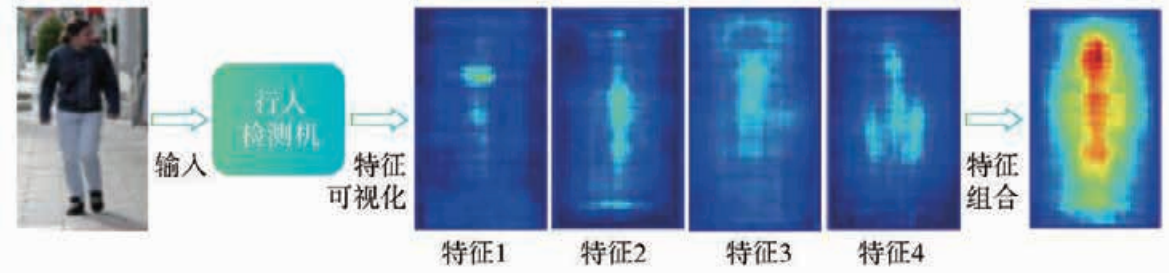


图 2-82　行人检测特征可视化

深度学习最新进展

随着深度学习在各个领域的广泛应用，多种多样的人工神经网络结构不断涌现，在相关科研和应用方面大放异彩。以下简要介绍深度学习尤其是深度卷积神经网络在物体识别、图像分割及图像生成等领域的最新进展。

1. 物体识别

2014 年，有研究者提出了基于候选区域的卷积神经网络 R-CNN（图 2-83）。不同于以往单阶段直接输出结果的识别网络，R-CNN 首先针对输入图片产生大量候选区域作为物体识别的中间结果，然后针对候选区域作进一步的识别。这种由粗到细的操作大大提高了物体识别的精度，使得 R-CNN 成为物体识别领域内的标志性深度神经网络架构。

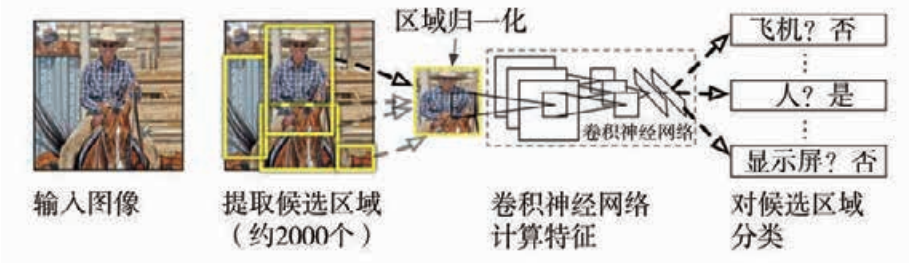


图 2-83　R-CNN：基于候选区域的卷积神经网络

2. 图像分割

图像分割一直是计算机视觉领域内的热点研究问题。2015 年有研究者提出使用全卷积网络（FCNN）进行图像中的物体分割。如图 2-84 所示，神经网络的前半部分由卷积层组成，随着网络的加深，分辨率逐渐降低。为了保证输出掩膜与输入图像的分辨率一致，神经网络的后半部分使用了上池化操作（增加特征分辨率）。这种设计方式使得 FCNN 的精度远超之前的模型。而且 FCNN 还引领了图像分割领域内的一股基于语义分割的浪潮。

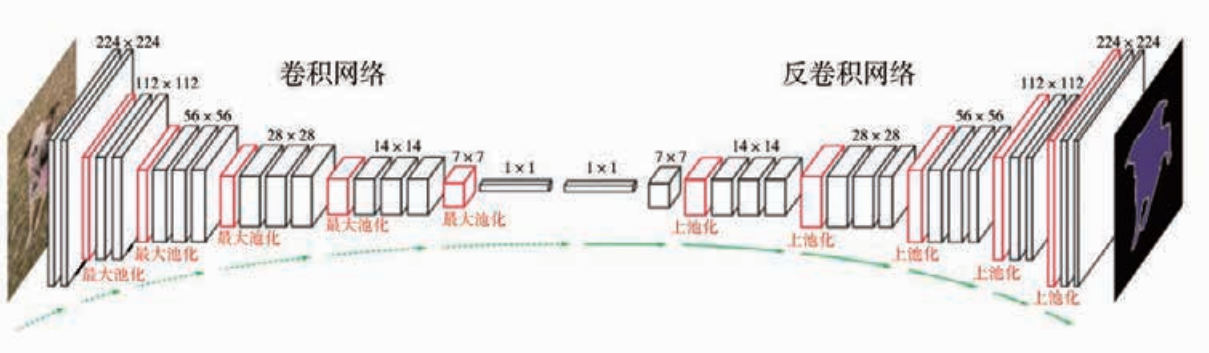


图 2-84　基于全卷积的图像分割网络

3. 图像生成

图像生成（图 2-85）是指利用一段随机噪声生成具有高级语义的自然图像，例如人脸、物体等。2014 年若干研究者提出的对抗生成网络（GAN）比较好地解决了这一问题。对抗生成网络利用一个判别网络判断由生成网络所生成图像的真实性，将它作为判断依据，不断训练生成网络，提高其生成效果。而相反的生成网络则尽可能迷惑判别网络，从而形成对抗。这种对抗生成的态势极大地提高了图像的生成质量，引发了图像生成领域内的一波

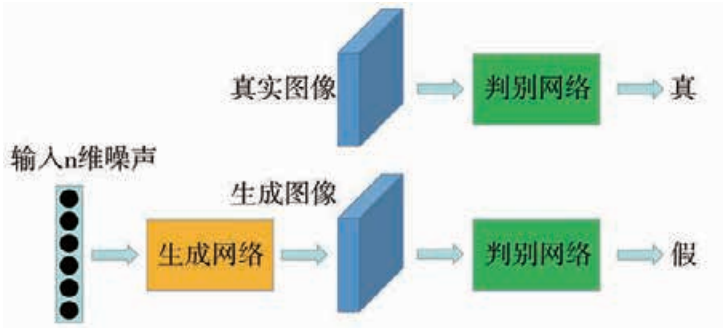


图 2-85　图像生成

研究热潮。

图 2-86 展示了对抗生成网络所生成的人脸示例。可以看出几乎达到了以假乱真的效果。

图 2-86　对抗生成网络所生成的人脸

拓展阅读

人工智能新高峰——AlphaGo

作为第一个战胜人类围棋职业棋手的人工智能程序，AlphaGo 在棋力上的飞速提升得益于很多人工智能技术，如深度学习、强化学习、蒙特卡洛树搜索等。

一、AlphaGo 的基本招式——蒙特卡洛树搜索

相比于上一个被计算机程序攻克的棋类运动——国际象棋，围棋的复杂度要高几十个数量级。如果穷举所有的可能性，国际象棋的复杂度在 35^{80} 种左右，而围棋在 250^{150} 种左右。在如此高复杂度的情况下，使用穷举所有可能性来与人对弈围棋显然是不可取的。AlphaGo 采用了一种有效的搜索算法，即蒙特卡洛树搜索算法（图 2-87），来提高学习落子策略的效率。

首先，我们定义第（j-1）次落子为 a_{j-1}，第（j-1）次落子后的棋面为 s_j。而蒙特卡洛树搜索的基本思想就是利用随机落子探索围棋规律。假设初始为空白棋面 s_0，初始所有落子 a 的分值都为 1，即（s_0,a）=1。经过不断随机落子直到第 n 次落子分出胜负后，我们把之前所有的落子策略（s_0,a_0）,…,（s_n,a_n）的分值都加 1，至此完成一次蒙特卡洛树搜索。把这个过程重复多次（比如 100 万次），逐渐地，落子胜率较高的策略相应分值就会比较高。在实际落子时，可以根据学习到的策略进行，于是得到一个具有一定棋力的算法。

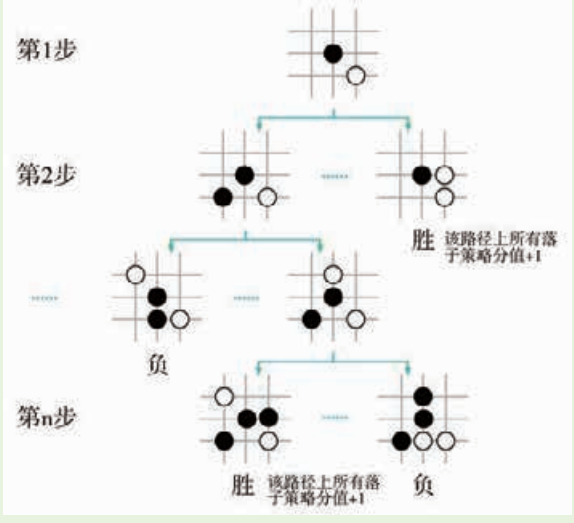


图 2-87　蒙特卡洛树搜索示意图

二、AlphaGo 的落子策略——深度学习

然而仅有“基本招式”还不足以支撑起 AlphaGo。在蒙特卡洛树搜索中，每一次落子都是随机生成的，这使得整个搜索的过程并不难，进而导致学习到的落子策略并不十分高明。相应地，人类棋手在下棋时，都会根据当前棋面判断局势，进而得出较优的落子策略。AlphaGo 团队由此联想到可使用深度神经网络模仿人类棋手的行为，代替蒙特卡洛树搜索过程中的随机落子，提高搜索效率。于是他们训练了一个 3000 万个训练样本监督下的策略网络。图 2-88 就展示了以当前棋面作为输入时，策略网络对不同落子位置的概率估计。概率越高代

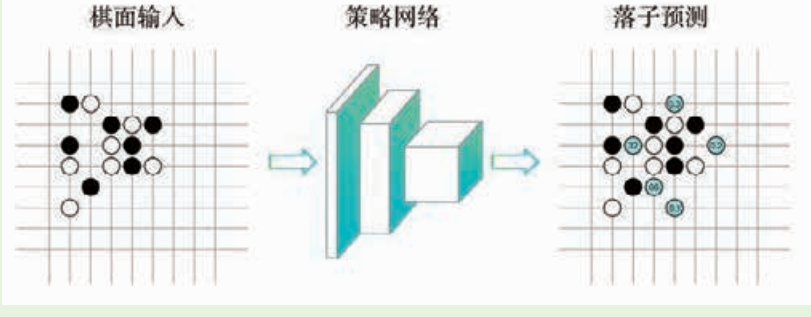


图 2-88　策略网络输出的落子概率

表人类棋手越有可能选择这个落子策略。在深度学习的帮助下，AlphaGo 已经可以战胜所有其他计算机，但是和人类职业选手还有一定的差距。

三、AlphaGo 的自我进化——强化学习

强化学习（Reinforcement Learning）是机器学习的一个分支。不同于监督学习，强化学习主要用于生成控制个体的行动策略。强化学习一般包括状态、动作、奖励、方案等要素。以围棋为例，个体是围棋程序，状态是棋面，奖励是输赢，动作是落子策略。AlphaGo 的强化学习过程是通过执行当前落子策略（动作）与棋面（状态）进行交互，根据输赢（奖励）不断改进落子策略的过程。由此可以发现，强化学习的基本思路是在对游戏完全没有了解的情况下，通过不断训练（进行多盘对弈，并获得做了动作后的分数反馈）来进行自我提升。

在强化学习的框架下，AlphaGo 团队设计了一个评价函数 v(s)。如图 2-89 所示，蓝色部分颜色越深，代表预测的落子胜率越高。此函数的功能是评估围棋局面的形势，以作为落子策略参考的一部分。使用 v(s) 可以让计算机在搜索的过程中不用走完全局（走完全盘耗时耗力，效率不高）即可判断胜负。这样可避免搜索到底，从效率（剪枝，优化算法时间复杂度）上进一步增加蒙特卡洛树搜索的威力。为提高效率，AlphaGo 团队还使用机器和机器对弈的方法来创造新的对局，也就是 AlphaGo 的左右互搏。

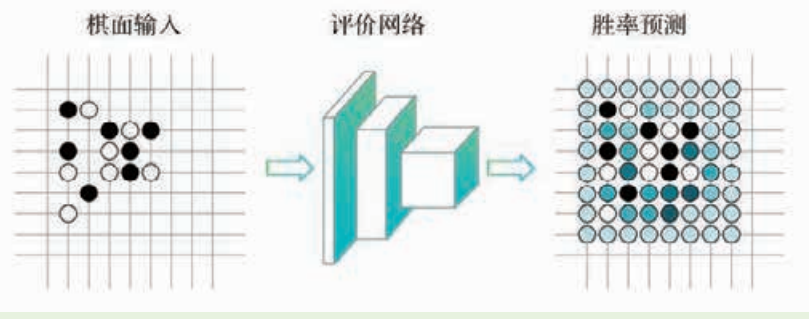


图 2-89　评价网络输出的落子胜率

在以上三大技术的帮助下，AlphaGo 在进行实际对弈时，首先利用蒙特卡洛树搜索算法预测棋局走势。在策略网络的帮助下，AlphaGo 不再随机预测落子，而是以更接近于人类棋手的思维推演棋局，以实现更加合理的棋面估计。同时在评价网络的帮助下，AlphaGo 可以更综合地考量具体每一个落子策略在当前棋面下的胜率，并把这一胜率也加入到棋局估计中。综合未来可能走势和落子策略，AlphaGo 可以根据当前棋面进行模拟推演，从而完成一步蒙特卡洛树搜索。不断重复这样的步骤直至分出胜负，则得到当前棋面下落子策略的分值分布。在实际落子时，AlphaGo 采用的是分值最高的那一个落子策略。至此，融合了蒙塔卡洛树搜索、深度学习和强化学习的 AlphaGo 不断进化，最终战胜了围棋职业高手，一战成名。

——摘译自 David Silver,et al.Mastering the game of go without human knowledge.2017

项目七

在车展中实现“车以类聚”

——探究无监督学习与聚类算法

某城市要举办一场大型车展，参展的各类车辆有上千辆。主办方希望根据车辆类型，将它们安排到不同的展区。然而，车辆种类繁多，包括轿车、卡车、工程车、消防车，以及一些叫不出名字、不知其功能的车辆。项目三中学到的方法可对车辆进行分类，但前提是要有分类信息。如何能在没有分类信息的条件下将车辆分到不同的展区呢？这就要用到人工智能无监督学习中的聚类算法来进行车辆聚类（图 2-90）。

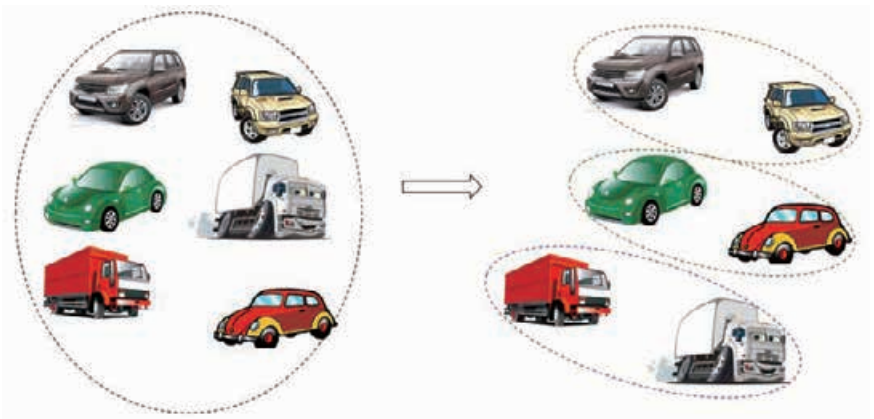


图 2-90　车辆聚类

项目学习目标

在本项目中，我们将探究运用无监督学习中的聚类算法将车展中的车辆安排到不同展区的基本思想和方法。

完成本项目学习，须能回答以下问题：

1. 无监督学习与监督学习的区别是什么？
2. 聚类与分类有什么不同？
3. k- 均值算法的思想方法是什么？
4. 如何度量聚类的好坏？

项目学习指引

我们之前学习的算法均是有监督学习，它的一个鲜明特征是通过给定的标签来学习从数据特征（如图像）到语义标签的映射关系。但在很多实际问题中，数据并没有语义标签，解决此类问题就要用到无监督学习。

无监督学习有很多技术方向，聚类是其中一个重要的方向。聚类的本质就是把特征相近的数据样本放到一起，达到“人以群分，物以类聚”的效果。将聚类应用在社交网络上，可以根据上网习惯特征，把行为习惯类似的用户归为同一组别，以便向同一组别的用户推荐相同的产品。

1. 认识无监督学习与聚类算法

（1）监督学习和无监督学习

机器学习分为监督学习（supervised Learning）和无监督学习（unsupervised Learning）。

参见 P90 知识链接“无监督学习”

① 监督学习

项目三中，智能车会根据数据库中的人脸来识别车主。在识别过程中，智能车将待识别对象与车主数据库中的标签（备注信息，说明是车主或不是车主）对象进行对比，从而判断输入的人脸是否属于车主。每次识别产生的明确答案（是车主或不是车主）又会成为该识别对象的标签，成为之后进行人脸识别分类的参考标准。

像这样需要根据类别的标注信息来进行分类的人工智能属于监督学习。每次识别后形成的答案（对象的标签）成为训练智能车的老师，教会智能车准确识别车主。

② 无监督学习

根据分类标签，我们可以很快找到一条最优直线，对目标对象进行分类。当数据量小的时候，先标注标签，然后再分类，分类很快就能完成。但当数据量巨大时，用人工标注方式来分类则非常繁琐，比如给成千上万的照片手动标注“是什么”的标签，不但工作量巨大，而且标注过程中还可能产生很多错误。因此，人在对物体进行分类的时候，有时不需要分类标签信息。比如说电影鉴赏，没有人告诉我们某部电影的类别，也不知道要把这部电影归于哪一类，但是在看完

很多电影之后，自然会把《宝莲灯》与《大闹天宫》归为一类，把《叶问》与《霍元甲》归为另一类。

同理，人工智能在处理海量数据时，如果通过预处理（加标签）使数据满足分类算法的要求很难实现，通常就采用无监督学习。这样人工智能可以不需要分类标签，而是在不断识别数据的过程中学会自己分类。这种不需要标注信息的学习过程称为无监督学习。

（2）分类与聚类

分类作为一种监督学习方法，要求必须事先明确知道各个类别的信息，并且保证所有待分类项都有一个类别与之对应。但是很多时候上述条件得不到满足，尤其是在处理海量数据的时候。如果通过预处理使数据满足分类算法的要求，代价非常大，那么就可以考虑使用聚类算法（clustering algorithm）。

参见 P90 知识链接“聚类算法”

聚类算法是一类比较典型的无监督学习算法。它有一个前提假设：同一个类的样本，特征应该相似。这是符合生活常识的，如恐怖片的情节、音乐都比较惊悚，喜剧片的情节、音乐都比较欢快。基于这个假设，可以根据样本间的相似度，将它们分到各个小组，并且让每个小组中的样本特征尽可能相似，让不同小组的样本特征尽可能不相似。从上面的描述可以看出，衡量样本间的相似度是聚类算法的基础。

2. 剖析 k- 均值聚类算法

（1）计算特征相似度

计算样本间的相似度实际上是计算样本特征值的相似度。计算特征值相似度通常分三步，如图 2–91 所示。

活　动

7.1 探究监督学习和无监督学习的区别。

（1）分小组查阅相关资料，详细了解监督学习和无监督学习的特点。

（2）根据生活经验，举生活中的案例说明监督学习和无监督学习。

（3）形成小组探究报告。

图 2-91　计算特征值相似度的步骤

采集到的特征经过数学建模后会形成特征值，特征值通常可以用坐标系中的坐标点表示。在数学中，两个坐标点的远近可以通过计算它们的距离来衡量。同理，特征值的相似度也可用特征值坐标点间的距离来代表。距离越小，特征值越接近，即特征值所代表的物体越相似。

聚类算法在采集到特征值后，会计算特征值间的距离，距离越小说明相似度越高，并以此为依据将样本分类。如图 2-92 所示，猩猩的特征值是（2，2），鸟的特征值是（1，0），猴子的特征值是（1，2）。从中可以看出，猴子与猩猩的特征值距离最小，因此得出猴子与猩猩相似度最高。

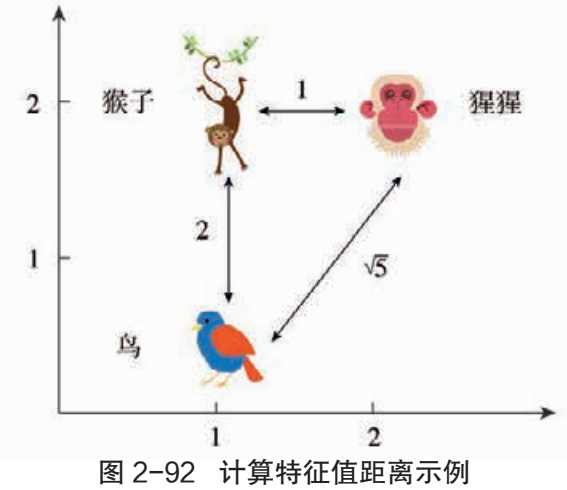


图 2-92　计算特征值距离示例

小贴士

参考项目三中关于如何采集特征及如何形成特征值的内容。

活　动

7.2 选择生活中的一些物品，如自行车、摩托车、汽车等，自己定义所需要统计的物品特征，然后计算它们的相似度，并验证是否与生活常识相符。查阅相关资料，查找还有哪些方式可以计算特征值相似度。

（2）k- 均值聚类算法的思想

k- 均值（k-means clustering）算法是典型的聚类算法，它可根据物体的特征值相似度对物体进行分组。

k- 均值算法的核心思想是：给定组别数 k，选 k 个组长，然后把所有待分组的样本分到与之距离最近的组长所在的组。组长被称为聚类中心（cluster center）。“均值”的涵义在于每个组长的特征值为该组所有成员的特征平均值。由于每个样本与自己所在组的“组长”距离最近（即最“相似”），可以预见同一组内的样本均比较相似。

参见 P90 知识链接“聚类中心”

这里有一个问题：每个组组长的特征值是该组样本特征值的平均值，这意味着每个组的样本特征值决定了组长的特征值。而每个样本又需要通过计算与组长的相似度（距离）才能被分组。这就好比“先有鸡还是先有蛋”，成了循环依赖了。像这种循环依赖的问题，需要首先确定一方的情况，才能打破僵局。

小贴士

针对“先有鸡还是先有蛋”这样的循环依赖问题，可选择每步固定一个变量，然后优化另一个变量，如此交替迭代优化，直到得到一个收敛的解。判断其收敛的依据是：两个变量在几次迭代中变化不大。

k- 均值聚类算法的核心思想在于“循环迭代”。首先随机给出每组的聚类中心；然后计算对应的最佳组别分配；接着反过来再计算最佳的每组聚类中心。依次循环迭代，直至最终的样本分组情况不再变化。

k- 均值聚类算法的步骤为：

第一步，随机地从所有样本中选取 k 个样本，作为每一个组的初始聚类中心。

第二步，将每一个样本分到与其距离最近的聚类中心所在的组，得到新的划分方式。

第三步，重新计算每组样本的聚类中心。

不断重复第二、三步，直到聚类中心和样本分组情况不再变化为止。

（3）探究 k- 均值聚类算法的迭代过程

假设车站有 9 辆车，如图 2-93 所示，第一次分组时，首先随机选取三辆车（带框的车）作为三个组长，并确定组名分别为：美猴王、猪八戒、沙和尚。然后计算每辆车与三个组长的距离，并将其分到距离最近的组长那一类。从图中可以看到，A 车与“猪八戒”的距离最近，因此被分到猪八戒的那一组，其他的车类似，于是得到第一次迭代分类的结果。

进行第二次迭代时，聚类中心不用随机选择，而是计算每个组中车辆的平均特征值，并以它作为聚类中心。

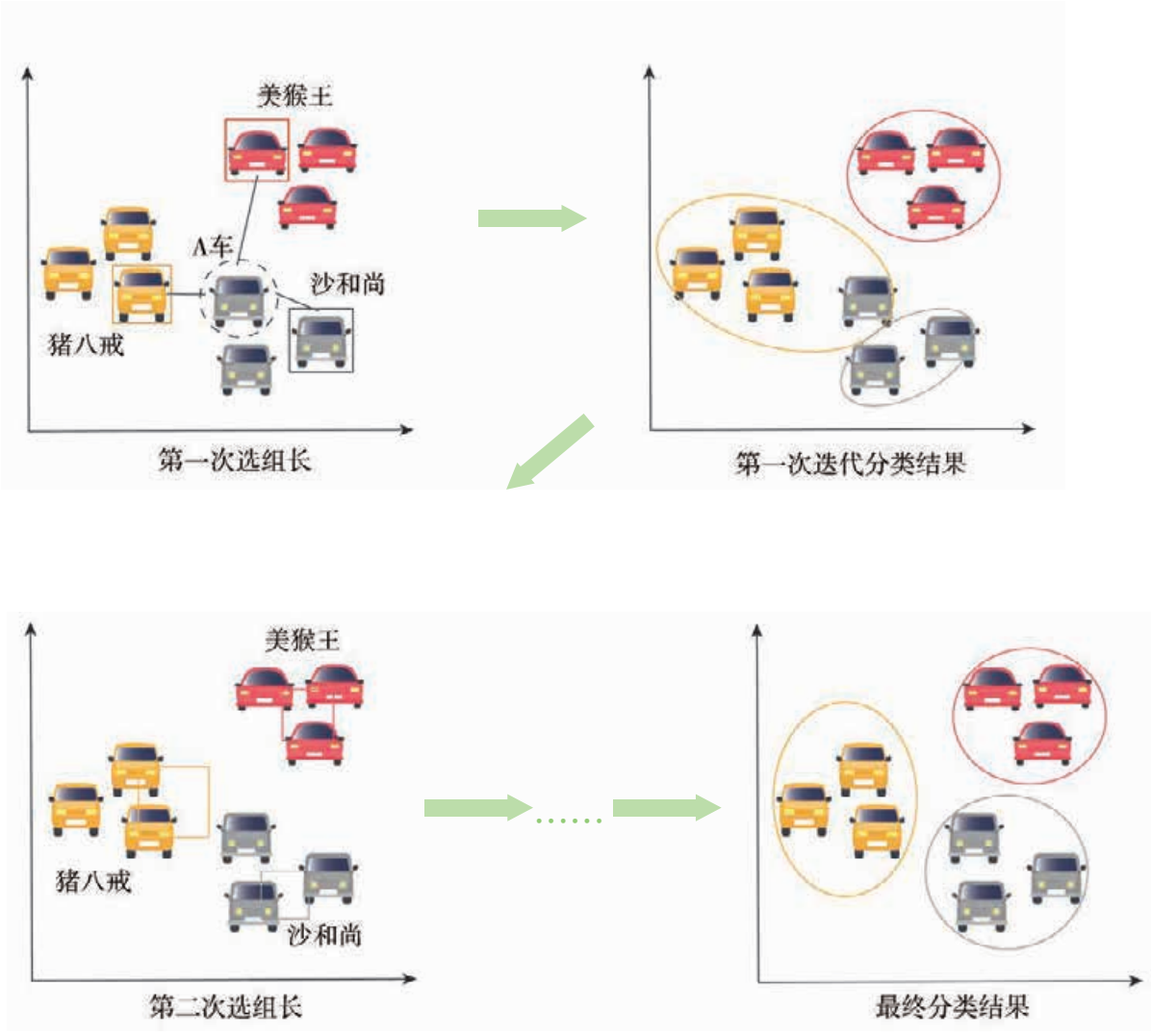


图 2-93　k－均值聚类算法的迭代过程

然后根据新聚类中心对车辆重新分组。重复这两个步骤，直到分组结果没有变化为止。

可以看出，有 n 个样本：$x_1,x_2,\cdots,x_n$，k– 均值聚类算法的任务是将这 n 个样本聚类成 k 类，这 k 个类各自的聚类中心分别为 $m_1,m_2,\cdots$，m_k。其迭代过程主要分为两个步骤：

第一步：选聚类中心。第一次迭代随机选择聚类中心；其他时候，计算每个类的平均特征值作为新的聚类中心。以第 k 类为例，假设该类中含有 n 个样本 $x_1,x_2,\cdots,x_n$，则其聚类中心更新计算公式如下：

$$m_k = \frac{1}{n}\sum_{i=1}^{n} x_i$$

第二步：为每个样本找与其最近的聚类中心。记 x_i 与 m_j $(1 \leqslant i \leqslant n,\ 1 \leqslant j \leqslant k)$ 的距离为 $d(x_i, m_j)$。在得到新的 k 个聚类中心之后，将每辆汽车分到与其距离最近（最相似）的聚类中心那一类。以样本 x_i 为例，应将它分配到与其距离最近的聚类中心 m_k 那一类，即 x_i 与 m_k 的距离 $d(x_i, m_k) = \min(d(x_i, m_i))$，i=1，2，…N。

活　动

7.3 体验 k- 均值聚类算法。

假设有一次车展活动，总共有 4 辆汽车参展。现在需要根据每辆车的特点，使用 k- 均值聚类算法将车辆分为两组。车辆的各项特征值如表 2-6 所示。

表 2-6　汽车特征统计表

汽车编号	大小	颜色	数学建模
A	5	4	A（5，4）
B	4	3	B（4，3）
C	4	2	C（4，2）
D	2	4	D（2，4）

（1）在下面的二维坐标系（图 2-94）中标注出代表 A、B、C、D 汽车的坐标点，并直观地对 4 辆汽车进行聚类。

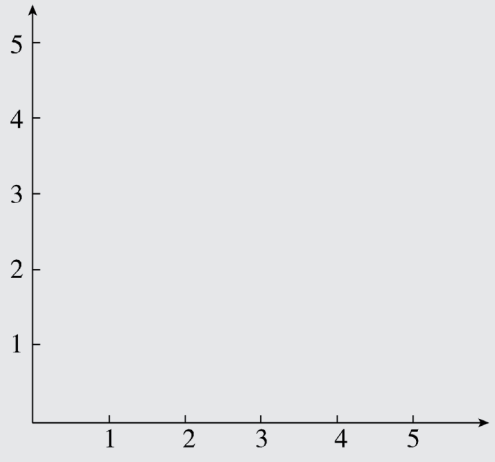


图 2-94　二维坐标系

（2）通过计算进行聚类。

第一次聚类计算：

步骤 1：随机选取聚类中心。不妨选择 A、C 作为聚类中心，并记为 m_1、m_2，则 m_1=（5，4），m_2=（4，2）。

步骤 2：计算每辆汽车与聚类中心的欧氏距离，并进行聚类［将 A、B 两点距离记为 d（A，B）］。

由计算可知，汽车 B 离_____组更近，汽车 D 离_____组更近，因此，汽车 B 应该分到_____组，汽车 D 应该分到_____组。故 m_1 组有_____车，m_2 组有_____车。

第二次聚类计算：

步骤 1：计算聚类中心。

计算 m_1 组的聚类中心的特征值，并记为 m_1=_____。

计算 m_2 组的聚类中心的特征值，并记为 m_2=_____。

步骤 2：计算每辆汽车与两个聚类中心的欧氏距离。

步骤 3：根据计算，对汽车重新进行聚类。

最终，m_1 组有_____车，m_2 组有_____车。

（3）通过编程进行聚类。

利用基于 Python 的机器学习工具 scikit-learn 中的 k- 均值聚类算法代码进行聚类，代码见学习资源包。

（4）对比（1）到（3）的结果，谈谈你对 k- 均值聚类算法的看法。

7.4 思考聚类中心的个数对分类结果会产生怎样的影响。

（4）无监督学习的应用

通过聚类，人工智能在没有数据类别标签的情况下，可使用数据本身特征的相似度来衡量数据是否属于同一类。

这种无监督学习算法不需要海量的数据标签，所以无监督学习不仅在聚类中有应用，而且在目标检测、图像生成等很多领域中都有重要应用。

知识链接

无监督学习

监督学习需要大量的数据标注，与之相反，无监督学习不需要数据的任何标注信息，仅根据数据本身的特征属性挖掘知识，完成任务。可以这样理解：一份试卷的参考答案就是标注信息，如果试卷有参考答案，对照参考答案做试卷就是有监督学习，如果试卷没有参考答案，那么做试卷就是无监督学习。无监督学习方法不仅在数据分类上有应用（聚类），在数据降维（典型算法为 PCA）等其他方面也有应用。

聚类算法

聚类算法是无监督学习的一个典型算法。对于无标签的数据，聚类算法可通过衡量特征值的相似度，将数据分到不同的类。除了 k- 均值聚类算法，聚类算法还有 DBSCAN、BIRCH、CURE 等聚类算法。

聚类中心

聚类中心就是每个类中的标准样本，也就是最符合这个类的特点的样本。不同的聚类算法对聚类中心的选定各不相同。k- 均值聚类算法是取该类所有样本特征值的平均值作为聚类中心。选定聚类中心之后，样本与聚类中心距离越小，样本和该类的标准样本就越相似。

拓展阅读

无监督学习神经网络模型的研究

从很多没有标注的信息中学习到知识是人类的智慧本能。例如，出于爱好，我们会阅读大量书籍，没有人告诉我们书里有什么，但是我们能够通过大量的阅读发现书本中的知识，比如作者的行文风格、思维方式。

2012 年，吴恩达（Andrew Ng）等人提出一个无监督学习神经网络模型，他们模仿人脑视觉皮层，将模型分成三个部分：

1. 特征获取：扫描图像的各个区域，将这些区域的内容特征“记录”下来。

2. 特征处理：图像中相邻区域的内容通常有一些重复，所以模型需要精简重复内容，选出一个区域里最具有代表性的内容特征。

3. 特征学习：从经过提炼精简的内容特征中学习知识，分析内容中的重要成分。设计好模型之后，让模型一直“观看”视频图像，并且不提供任何辅助信息。三天后，测试发现，这个网络模型成功地学习到了如何分辨一幅图像中是否有人脸、人的身体部分或猫脸的能力，如图 2-95 所示。这个发现展示了人工神经网络自主学习知识的能力，极大地推动了无监督学习在人工神经网络中的研究进展。

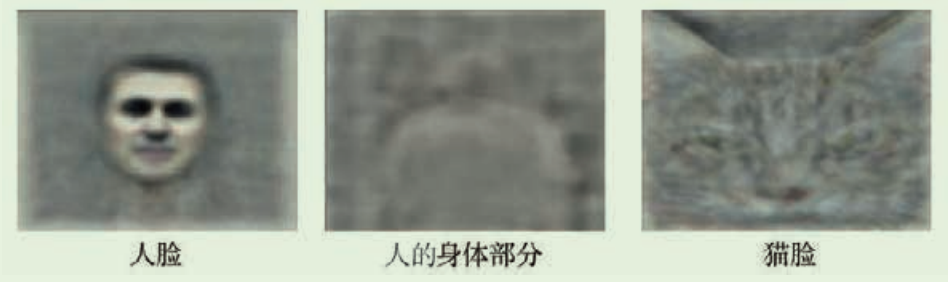


图 2-95　网络模型分辨结果

——摘译自 Quoc V.Le,et al.Building high-level features using large scale unsupervised learning.2013

单元挑战　用 SVM 算法及深度学习给图像分类

一、项目任务

SVM 算法和深度学习已经被广泛地应用在图像分类任务等方面。如图 2–96 所示，选取四类图像，每类图像有 15 幅。请大家使用前 4 行的图像进行训练，使用最后一行的图像进行测试。通过实验来探索 SVM 算法与深度学习在图像分类中的应用。

图 2–96　四类图像

二、项目指引

1. 将每类图像的前四行（12 幅 ×4 类）作为训练集进行训练，将最后一行（3 幅 ×4 类）作为测试集进行测试。

2. 在模型训练中，设置 10 组不同的学习率和步长，对比学习率和步长对损失函数收敛速度的影响以及对最终测试分类结果的影响。

相关代码和模型见配套资源。

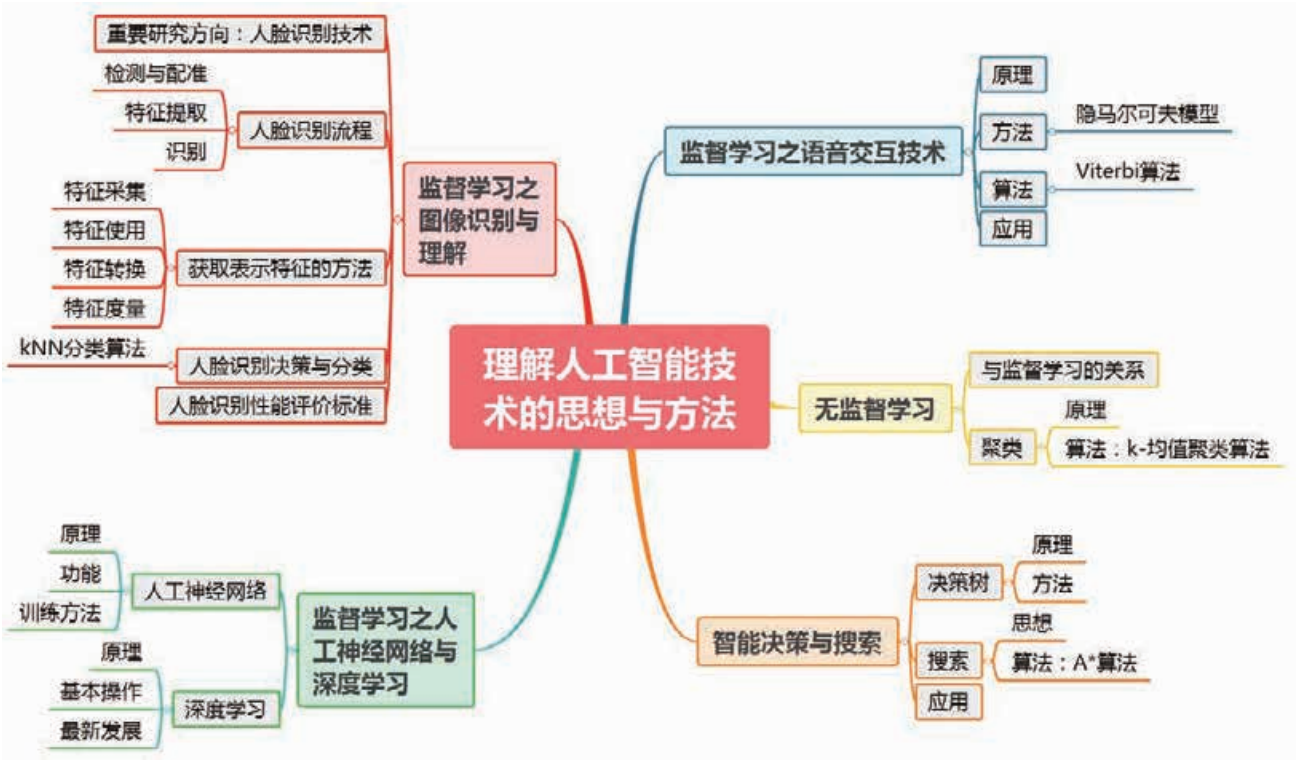
三、交流评价与展示

1. 分别取出在测试集测试中用 SVM 算法和深度学习取得最佳分类结果的模型，然后分析对比这两组模型在精度、速度、内存消耗量和稳定性上的区别。

2. 每个小组展示交流得到的分类结果。

单元小结

一、主要内容梳理



二、单元评价

评价内容	达成情况
能说出图像识别技术的主要应用（A、R）	
能说出人脸识别的流程（A、T）	
理解特征的意义、获取方法和表示方法（A、T）	
能说出人脸识别的常见算法的思想（A、T）	
知道语音识别的流程（A、T）	
能说出隐马尔可夫模型对语音识别的作用（A、T）	
能说出语音识别领域中主要算法的思想（A、T）	
能说出决策树和启发式搜索在路径规划中的作用（A、T）	
能说出启发式搜索的基本思想（A、T）	

知道人工神经网络的原理（A、T）	
理解人工神经网络与深度学习的关系（A、T）	
知道深度学习的基本操作：卷积操作和池化操作（A、T）	
理解深度学习目前遇到的难题（A、T）	
能举例说明监督学习和无监督学习的区别（A、T）	
理解无监督学习中的 k- 均值聚类算法的思想方法（A、T）	

说明：A—信息意识，T—计算思维，I—数字化学习与创新，R—信息社会责任

第三单元

开发简单人工智能系统

如何进行人工智能系统开发呢？一般来说，人工智能系统由硬件部分和软件部分组成，开发系统其实就是对硬件部分进行组装，并通过编程进行软件设计，最终形成人工智能系统。

目前有很多开源开发平台和开源人工智能应用框架帮助开发者开发简单的人工智能系统。树莓派（Raspberry Pi，简写为 RPi 或 RasPi / RPI）是一种为学习计算机编程而设计，只有信用卡大小的微型计算机。树莓派结合基于 Python 语言的开源应用框架 / 平台，能帮助开发者相对容易地开发人工智能系统。

在本单元中，我们将使用树莓派，以及基于 Python 语言的开源视觉智能应用框架 OpenCV，搭建智能小车，体验简单的人工智能系统开发。

单元挑战

设计智能车自动避障系统

学习目标

◆ 利用开源人工智能应用框架，搭建简单的人工智能应用模块，并能根据实际需要配置适当的环境、参数及自然交互方式等。

项目八

搭建可“刷脸”启动的循迹智能车
——设计简单的人工智能系统

循迹机器人是一种能够自动按照给定的路线进行移动的机器人，它是一个运用传感器、信号处理、自动控制等技术来实现路面探测、障碍检测、信息反馈和自动行驶的技术综合体。常见的循迹方法是通过光电传感器（如红外 / 超声波传感器、摄像头）探测地面上不同标识的模式（如色彩模式——黑色和白色、红色和绿色，形状模式——圆形、正方形、三角形）来获得引导信息，修正机器人的运动路径。循迹机器人在物流等领域已得到了应用，例如自动化生产线的物料配送机器人、仓库搬运机器人（图 3-1）等。

而利用人脸识别技术和循迹技术，便可设计开发出可通过“刷脸”启动并且可以按照路线自动行驶的智能车。

图 3-1　仓库搬运机器人

项目学习目标

在本项目中，我们尝试利用树莓派和 Python 语言的算法库，搭建一辆可“刷脸”启动的循迹智能车，体验搭建简单的人工智能系统的过程与方法。

完成本项目学习，须能回答以下问题：

1. 智能车系统由哪些部分构成？
2. 如何使用 Python 的开源人工智能库？
3. 如何配置树莓派环境？
4. 如何进行项目分解？

项目学习指引

搭建人工智能系统，首先要明确项目的目标，特别是项目对“智能”的要求。本项目要求搭建的智能车的智能要求主要有两个：①“刷脸”启动；②循迹，即沿着地面的信号标记前进。

明确项目目标后，可以从可行性的角度，对智能车的设计形成初步构想，做好相应的工具、套材准备和知识准备。

1. 进行总体设计，确定基本开发方案

（1）分析智能车系统

可“刷脸”启动的循迹智能车主要具有两个功能：“刷脸”启动和循迹前进。本项目可以分为两个子项目：实现人脸识别模块和实现自动循迹模块。

如表 3-1 所示，人工智能系统通常由感知机构、决策机构和执行机构组成。智能车的感知机构可采用摄像头来接收环境信息；决策机构可采用树莓派开发板（图 3-2）来实现算法，制定行为决策；执行机构可采用驱动控制板来控制电机转动，从而带动车轮，使智能车运动。

表 3-1　树莓派智能车的机构组成

感知机构	决策机构	执行机构
摄像头	树莓派开发板	驱动控制板、电机与车轮

图 3-2　树莓派

智能车的搭建如图 3-3 所示。

（2）确定开发方案

制订智能车系统的开发方案是决定智能车开发方式的重要一步。以树莓派智能车系统环境搭建的开发方案为例。这个开发方案使用的基本硬件为树莓派开发板 3 代 B 型、树莓派官方摄像头以及驱动控制板；基本操作系统为 Raspbian

小贴士

一般来说，一个大项目很难一次性地完成，因此人们常将一个大项目分解为多个子项目，然后“各个击破”，最终完成整个大项目。

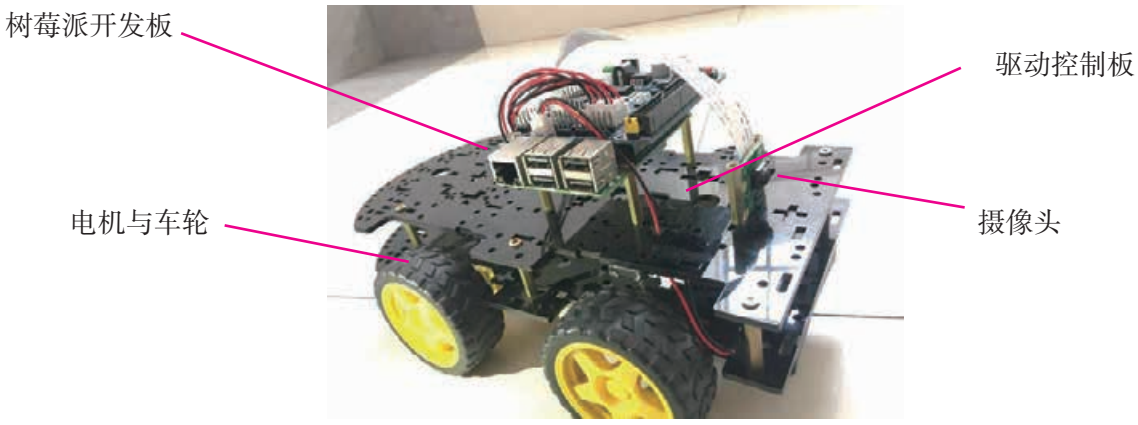


图 3-3　搭载树莓派的智能车

系统，系统的编程语言为 Python，并且引用了 Dlib、scikit-learn、OpenCV 等开源的人工智能库，如图 3-4 所示。

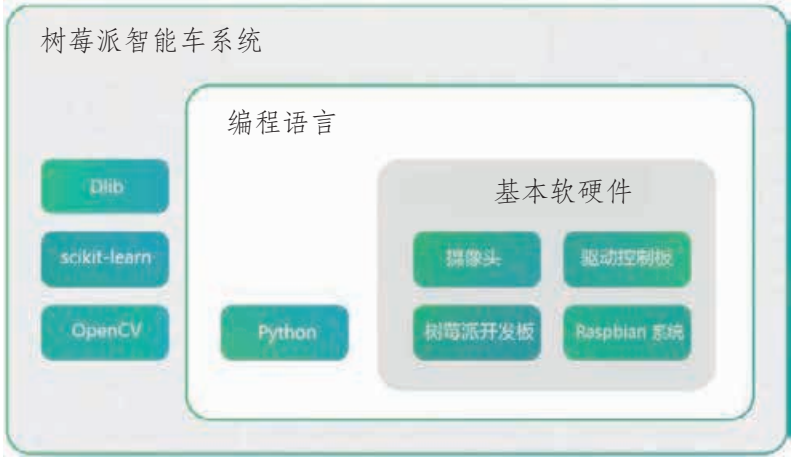


图 3-4　智能车系统环境搭建

可“刷脸”启动的循迹智能车的工作流程如图 3-5 所示。树莓派启动之后，系统进入人脸识别模块，并持续读取摄像头采集的人脸图像，通过人脸识别算法判断是否与已登记的人脸一致。如果一致，则识别成功，系统会进入自动循迹模块。在自动循迹模块中，系统持续调取摄像头采集的环境图像，检测出地面标识的位置并进行循迹。如果检测到特定的停止标识，则停止循迹，流程结束。

2. 设计人脸识别启动系统

（1）初步设计

智能车需具备简易的人脸识别启动系统。该系统可这样设计：通过人脸识别算法对摄像头采集的包含正面脸部的照片进行处理，与数据库中预先存储的照片比较，最后在屏幕上显示识别结果。如果识别成功，智能车启动。

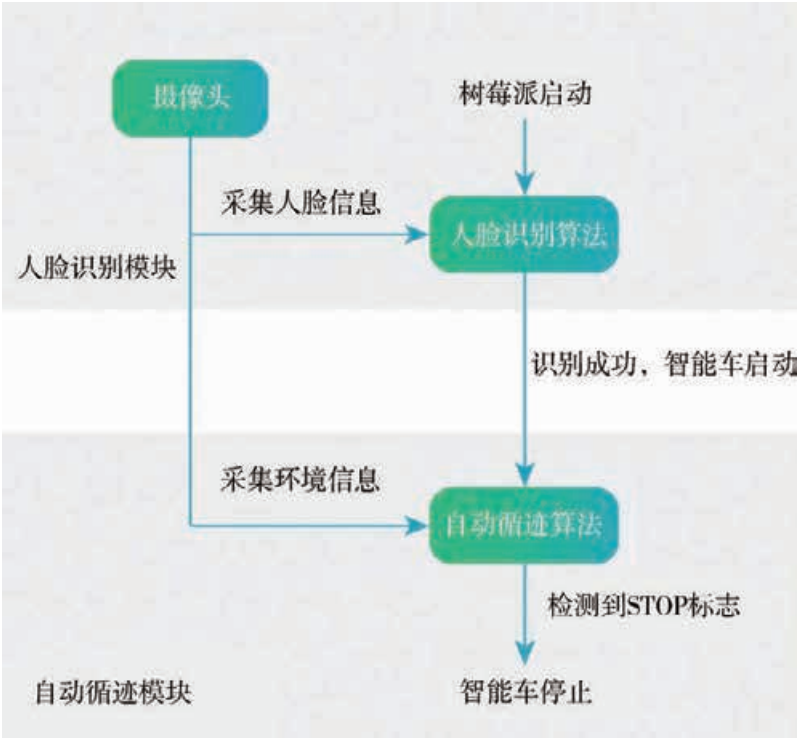


图 3-5　树莓派智能车系统实验流程图

智能车的人脸识别启动系统可以使用树莓派开发板实现，过程可分为以下三步：

① 建立一个保存有 N 张不同人脸的数据库；

② 在树莓派开发板上实现一个基于深度学习的人脸识别算法；

③ 对当前摄像头采集到的照片进行识别，得到识别结果。

整个流程概况如图 3-6 所示。

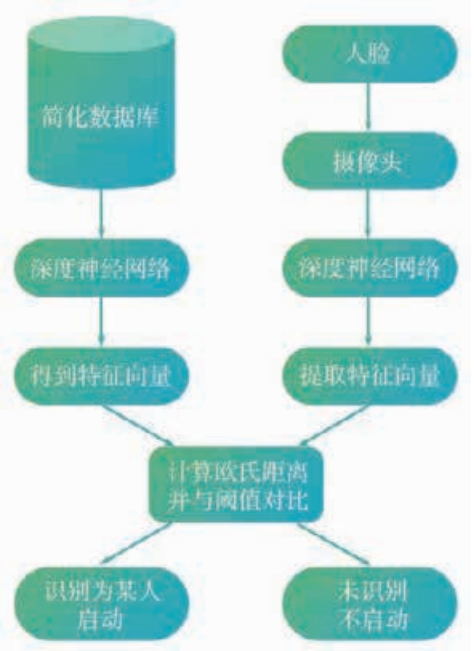


图 3-6　人脸识别启动系统流程图

建立数据库时，若不方便创建专业数据库，则可创建一个名为“img”的文件夹，将需要登记的人脸照片放入这个文件夹中。之后可在程序代码中直接读取这些图片。

建立数据库后，可采用残差网络（ResNet）来提取人脸照片的特征向量，如图 3-7 所示。输入一张包含人脸的照片，经过 ResNet 处理后得到一个 128 维的特征向量，之后用特征向量之间的距离来度量人脸的相似程度。

小贴士

残差网络是一种性能优越的深度卷积神经网络。它的出现使上百甚至上千层的神经网络的训练变得容易。

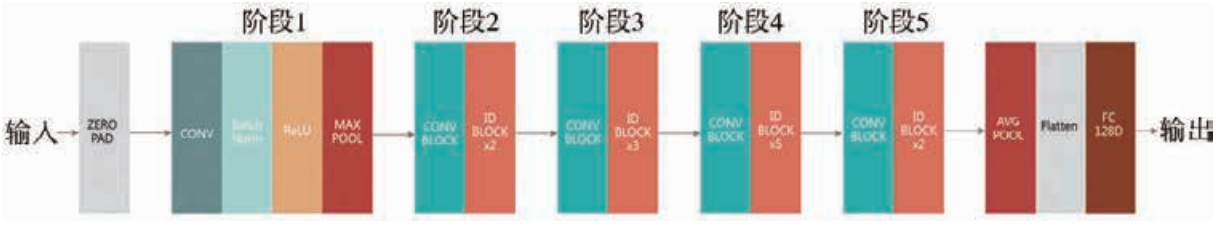


图 3-7　利用残差网络提取特征向量

（2）配置开源人脸识别库

要在树莓派上实现人脸识别，简单快速的方法是使用基于 Python 的 face_recognition。该库使用了 Dlib 工具库的人脸检测和识别技术，具有很高的人脸识别准确率。

小贴士

Dlib 工具库是基于 C++ 的开源工具包，经过编译后可以在 Python 中直接调用，其中包含了基于 68 个关键点的人脸检测器以及已经训练好的 ResNet 人脸识别模型。

参见 P106 知识链接“安装编译 Dlib 工具库”

活　动

8.1 完成以下任务，实现人脸识别。

（1）根据知识链接中关于配置人脸识别库的说明配置人脸识别库。

（2）参考配套资源中关于实现人脸识别功能的代码，完成人脸识别的编程。

（3）创建并使用人脸数据库。首先使用树莓派摄像头给全班每位同学各拍一张照片，并将照片放入数据库文件夹中。人脸识别时，让摄像头同时捕捉多位同学的脸，观察多人识别的结果与单人识别的结果有什么不同。

（4）尝试做不同的表情，或稍微改变一下外形（戴上帽子、摘下眼镜等），或采用不同的光照条件，观察系统能否正确识别自己的脸，并分析其中的原因。

（5）根据测试思考：在人脸识别时，对于非真实的人脸会得到什么样的识别结果呢？测试参考：先将某位同学的照片放入数据库文件夹中，再将他的另外一张照片作为目标放在摄像头前，观察此时的识别结果，并思考其中的原因。

参见 P107 知识链接“实现人脸识别的代码”

3. 设计智能车循迹系统

（1）初步设计

智能车循迹系统的运行逻辑如图 3-8 所示。可以看到，智能车循迹系统由三个子系统构成，分别是环境感知系统、决策规划系统和控制执行系统。环境感知系统收集外界的信息，包括行人、车辆、障碍物等。这些信息交由决策规划系统进行处理，以规划出最佳的车辆行驶路线，然后向控制执

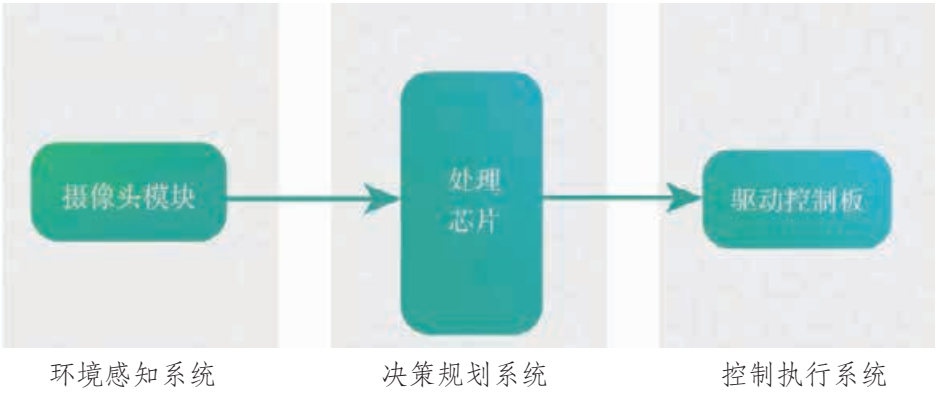


图 3-8　循迹系统的运行逻辑

行系统发送控制信号。最后控制执行系统驱动电机执行相应的动作。

以绿色标志代表可以通行，红色标志代表不能通行，智能车的循迹过程如图 3-9 所示。从上至下三个对话框分别代表环境感知、决策规划和控制执行三个阶段智能车的思维逻辑。

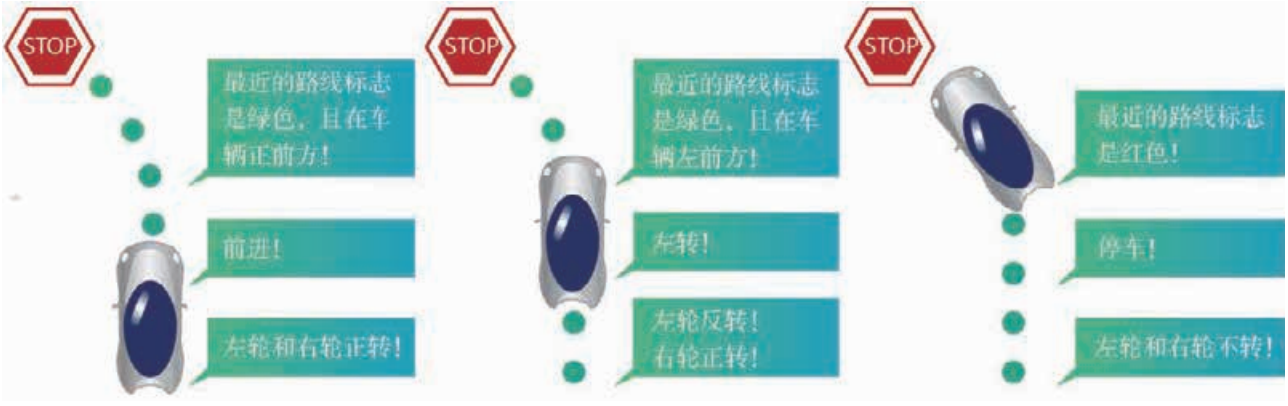


图 3-9　智能车的思维逻辑

（2）理解环境感知步骤——通过 k- 均值聚类算法获得路线标志的位置和颜色

在这一步骤中，需要找到离智能车最近的路线标志在摄像头中的位置和相应的颜色。具体过程是：

① 通过摄像头获取智能车前方的环境图像。

② 处理图像，找出所有可能在标志区域的像素点。

③ 用 k- 均值聚类算法求出上述像素点的聚类中心。

④ 找到离智能车最近的聚类中心，并判断其颜色。

活　动

8.2 通过编程实现环境感知。

（1）阅读并理解以下代码。

```
# 获取摄像头图像
image = frame.array
# 将图像转换为 HSV 格式
hsv = cv2.cvtColor(image, cv2.COLOR_BGR2HSV)
# 提取出所有颜色为绿色或红色的像素点
data = get_green_or_red_pixels(hsv)
# 对所有红色或绿色的像素点进行聚类。这里使用了 MiniBatchkmeans
# 损失少量精度换取比标准 kmeans 更快的速度
kmeans = MiniBatchKMeans(n_clusters = 2, random_state = 0, n_init = 1).fit(data)
# 提取出所有聚类中心
cc = kmeans.cluster_centers_
# 找出离摄像头最近的聚类中心的坐标，以（h， w）表示
(h, w) = find_closest_center(cc)
# 提取该聚类中心的颜色
color = get_color(hsv[h][w])
# 调用决策规划部分的代码，决定小车下一步的行动
# 根据聚类中心横坐标 w、图像宽度 width 和聚类中心颜色 color 决定行动指令
command = plan(w, width, color)
# 根据指令调用控制执行部分的驱动函数，这里设定转速为 20，每次移动 0.5 秒
move(command)
```

（2）执行上述代码，观察智能车感知环境的能力。

（3）尝试改变聚类中心 n_clusters 的值。观察当 n_clusters 的值分别为 1、2 和 3 时智能车的循迹情况，并分析原因。

（3）设计决策规划——根据路线标志的位置和状态规划行动指令

通过环境感知部分，能够得到最近的路线标志在智

能车摄像头所采集图像中的水平位置及颜色。这里的图像信息决定了下一步的行动。

如果标志为红色，那就停车。如果不是红色，则根据聚类中心的坐标判断是应左转、直行还是右转。当坐标在左半屏幕时，发出左转指令；在右半屏幕时，发出右转指令；在屏幕中间时，发出直行指令。为了防止频繁的左转和右转，可以设置阈值界定左转、右转和直行的允许范围。

指令一旦发出，智能车执行机构将会立即根据指令执行相应的动作。

活　动

8.3 通过编程实现决策规划。

（1）阅读并理解以下决策规划的部分代码。

```
# 决策规划函数。参数 position_w 表示聚类中心的横坐标，width 表示图像的宽度
#color 表示聚类中心的颜色。
def plan(position_w, width, color):
# 将中间 20% 屏幕宽度的区域作为直行区域
    ratio = 0.2
    forward_width = width * ratio
    if color == "red":
        command = "stop"
    elif position_w<= (width – forward_width)/2:
        command = "left"
    elif position_w>= (width + forward_width)/2:
        command = "right"
    else:
        command = "up"
    return command
```

（2）若改变 ratio 的值，请思考智能车的运行会有什么改变。执行上述代码，根据实际情况选择一个最合适的 ratio 值。

（4）设计控制执行步骤——根据指令向驱动控制板发送控制信号

智能车基本的行动包括前进、左转、右转、后退。要实现这些功能，需要了解模型车的转向方式。

一般情况下，智能车的变向采用差速转向的方式：车辆通过控制左右两个驱动轮的正反转和转速实现转向。驱动轮转速不同时，即使没有像汽车那样的转向轮，也可以实现左转和右转动作。通过调节左右驱动轮正反转和转速，可以组合出不同的车辆行驶方式。

表 3-2　差速转向驱动轮转动与运动状态的关系

	左驱动轮正转	左驱动轮反转	左驱动轮不转
右驱动轮正转	前进	左转	左转
右驱动轮反转	右转	后退	\
右驱动轮不转	右转	\	停止

小贴士

注意：如果采用其他型号的驱动控制板，对应的接口和控制方式可能有所不同。请参考所用驱动板的说明书。

树莓派可采用 L298N 等驱动控制板进行差速转向控制。具体的原理：树莓派将驱动轮是否转动以及转速信息通过 GPIO 接口传递给驱动控制板，驱动控制板再根据此信息对电机进行供电和驱动。其中，采用高低电平来表示驱动轮是否转动，采用 PWM 脉冲波表示转速。

表 3-3　L298N 驱动板控制方式

直流电机	旋转方式	IN1	IN2	IN3	IN4
M1	正转	高电平	低电平	\	\
	反转	低电平	高电平	\	\
	停止	低电平	低电平	\	\
M2	正转	\	\	高电平	低电平
	反转	\	\	低电平	高电平
	停止	\	\	低电平	低电平

活　动

8.4 通过编程实现控制执行。

（1）阅读并理解以下控制执行的部分代码。

```
# 设定运动速度和运动时间，根据指令调用底层驱动函数 stop，left，right，up
def move(command):
    speed = 20
    time = 0.5
if command == "stop":
    stop()
if command == "left":
    left(speed, time)
if command == "right":
    right(speed, time)
if command == "up":
    up(speed, time)
```

（2）执行上述代码，观察智能车的运动情况。

（3）尝试改变 move 函数中转速 speed 和运动时间 time 的值，观察智能车的运动情况并分析原因。

4. 测试智能车

智能车完成后，必须对其进行测试。图 3-10 是已登记和未登记的两个人在摄像头前的识别结果。可以看到，对于已登记者，画面用红色方框标记出人脸的位置，并显示人名，而对于未登记者，则显示“Unknown Person”。

对已登记者的识别结果

对未登记者的识别结果

图 3-10　测试识别结果

以下两张照片是智能车循迹模块的实验结果图。在图 3–11 中，智能车检测到最近的绿色标识并沿路线进行移动。在图 3–12 中，智能车检测到红色标识，于是停车。

图 3–11　沿路线运动

图 3–12　停车

活　动

8.5 以小组为单位测试智能车，看看它是否按设计方案运行。记录出现的问题，并进行调整。调整完成后向全班同学展示智能车。

8.6 思考并完成以下任务。

（1）如果互换前进和停止标识的颜色，那应该对原来的代码进行怎样的修改才能够实现循迹的功能？尝试修改代码并实现这个功能。

（2）标识的形状对循迹的效果有影响吗？和同学们合作，分别用圆形、正方形、三角形和长方形的标识进行循迹，并观察效果。

知识链接

安装编译 Dlib 工具库

由于树莓派系统的默认交换空间（可以理解为内存）只有 100MB，不足以运行 Dlib 的人脸识别算法，所以首先需要把它的交换空间增加到 1G。步骤如下：

1. 打开设置交换空间的文件 /etc/dphys-swapfile，命令行输入：

```
sudonano /etc/dphys-swapfile
```

将 dphys-swapfile 文件中的 CONF_SWAPSIZE=100 改为 CONF_SWAPSIZE=1024。

2. 更新文件之后重新启动交换服务：

```
sudo /etc/init.d/dphys-swapfile stop
sudo /etc/init.d/dphys-swapfile start
```

3. 确认交换空间是否增加：

```
free -m
```

4. 从 Dlib 官网下载安装包，解压后进入 Dlib 目录，进行 Dlib 库的安装和编译：

```
cd dlib-19.9
sudo python setup.py install
```

5. 安装 face_recognition 库：

```
sudo pip install face_recognition
```

经过以上操作，即完成了人脸识别库的配置。

实现人脸识别的代码

1. 访问摄像头资源，并设置分辨率、帧率等：

```
# 访问摄像头
camera = picamera.PiCamera()
# 设置分辨率和帧率
camera.resolution = (320, 240)
camera.framerate = 30
```

2. 根据前面提到的用文件夹作为简化的数据库的方法，读取“img”文件夹下所有的人脸照片（文件格式为 jpg），并将 ResNet 提取的特征向量存储到内存中，以供后面比对：

```
# 创建名字 – 特征向量对字典
face_db = {}
# 以相对路径的方式获取“img”文件夹下所有的照片
imgs = os.listdir('./img')
for im in imgs:
    im_path = os.path.join('./img',im)
    # 加载“img”文件夹下的所有照片
    ima = face_recognition.load_image_file(im_path)
    # 检测照片中人脸的位置
    face_locations = face_recognition.face_locations(ima)
    # 获取人脸的特征向量
    image_face_encoding=face_recognition.face_encodings(ima,face_locations)[0]
    # 将照片的文件名作为人名（截取 .jpg 前面的部分）
    face_db[im[:-4]] = image_face_encoding
```

3. 主函数部分，调用face_recognition库函数，检测人脸位置并进行人脸识别，代码如下：

```
rawCapture=np.empty((resY*resX*3),dtype=np.uint8)
# 摄像头抓取一帧
camera.capture(rawCapture,format='bgr')
image=rawCapture.reshape(resY,resX,3)
# 人脸位置检测
face_locations = face_recognition.face_locations(image)
# 对检测到的人脸进行特征提取
face_encodings = face_recognition.face_encodings(image, face_locations)
# 与数据库中的人脸进行比对和识别
for face_encoding in face_encodings:
# 比对人脸，如果两张人脸的不相似程度小于阈值 threshold 则识别为已知的某个人，
# 否则认为是一个未登记的人
    threshold = 0.6
    match = face_recognition.compare_faces(face_db.values(),face_encoding,threshold)
```

单元挑战　设计智能车避障系统

一、项目任务

通过本单元的学习，查找资料，设计智能车避障系统。

二、项目指引

1. 布置场景，将停止标识看作障碍物随机摆放在地上。
2. 在智能车循迹系统的基础上，让智能车在遇到障碍物时能够绕开。

（1）分析智能车绕开障碍物的原理。

原理：

（2）设计避障算法。

避障算法：

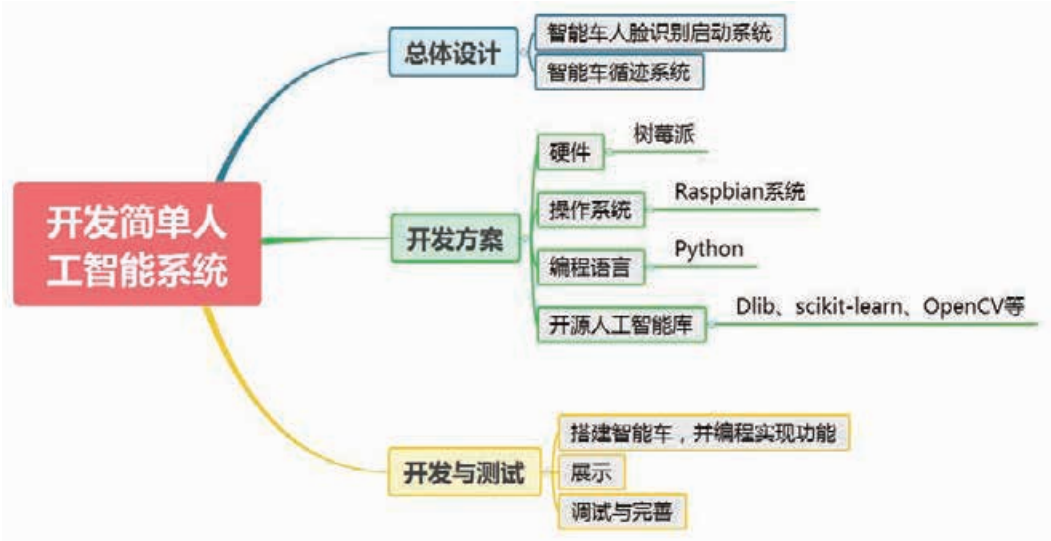
（3）在智能车循迹系统代码的基础上，编写实现避障的算法。

三、交流评价与展示

1. 各小组展示本组的智能车避障系统，讲解其避障原理。
2. 小组之间交流各自使用的方法，加深对于智能车系统的理解。

单元小结

一、主要内容梳理



二、单元评价

评价内容	达成情况
能说出简单人工智能系统的开发流程（A、T）	
能说出人工智能应用框架的作用（A、T、R）	
能说出开源硬件在人工智能系统中的作用（A、R）	
能说出项目分解的重要性（A、T、R）	
能配置开源人脸识别库（A、T）	
能对人工智能系统进行测试，找出问题并予以解决（A、T、R）	

说明：A—信息意识，T—计算思维，I—数字化学习与创新，R—信息社会责任

第四单元

推动人工智能健康发展

人工智能已经深入到社会生活的方方面面。一方面，随着相关技术的不断发展，人工智能在工业、医疗、社交、交通、金融等领域得到广泛应用，深刻地影响了人们的生活；另一方面，人工智能也给人类社会带来很大的冲击，例如，人工智能未来或许会取代人类的许多工作，导致大面积失业。那么在发展人工智能技术的同时，如何应对人工智能带来的社会问题，如信息安全问题和伦理道德问题，是我们需要面对的挑战。

在本单元中，我们将辩证地探讨人工智能带来的信息安全问题及伦理道德问题，了解保障人工智能安全可控和应对相应伦理道德问题的基本方法，展望人工智能的未来发展趋势。

单元挑战

设计无人驾驶时代的交通准则

学习目标

- 了解社会智能化所带来的伦理及安全问题。
- 知道信息系统安全的基本方法和措施，增强安全防护意识和责任感。
- 辩证地认识人工智能对人类社会发展的巨大价值和潜在威胁。
- 自觉维护和遵守人工智能社会化应用的规范与法规。

项目九

认识人工智能的巨大价值和潜在威胁

——辩证看待人工智能

无人驾驶作为人工智能的典型应用之一，一直备受关注。然而，时不时发生的无人驾驶引发人类伤亡的交通事故又提醒人们思考：人工智能安全吗？其实人工智能系统作为一种信息系统，确实存在其本身固有的安全问题，比如数据泄露、执行错误等。但人工智能系统又不同于一般的信息系统，它有“智能”，具备“学习”能力。这种能力是否需要受到人类的限制？这种能力不断发展会不会给人类生存带来威胁？人工智能时代（图4-1），这些问题更值得我们思考。

图 4-1　人工智能时代

项目学习目标

在本项目中，我们将一起分析人工智能的应用现状，探讨其中的伦理、安全问题，并畅想人工智能的发展前景。

完成本项目学习，须能回答以下问题：

1. 当前人工智能主要有哪些代表性应用？
2. 有哪些基本方法和措施来确保信息系统安全？
3. 如何辩证看待人工智能带来的极大便利和潜在威胁？
4. 人工智能的未来发展趋势是什么？

项目学习指引

生活中有很多人工智能应用产品，比如智能音箱，你只要说出想听的歌曲，它就能播放。甚至有时我们并未察觉人工智能的存在，比如智能新闻推送，它能根据我们先前在网上查看的新闻、商品等，主动推送相关的信息。这些人工智能技术的推广应用使得社会逐渐智能化。

1. 了解人工智能技术的应用现状

目前，我国正在努力推动人工智能的建设：一方面，大力构建互联网基础设施，包括计算中心和数据中心等；另一方面积，极引导相应公司建立人工智能平台，进行人工智能实践，如智能驾驶、智能城市、智能医疗和智能语音。

智能驾驶：人工智能技术可以辅助司机进行智能驾驶。从理论上讲，智能驾驶可实现无人化。智能驾驶将是未来解决交通拥堵的重要技术，能大大提升交通效率。百度公司开发的自动驾驶平台提供了一套完整的软硬件和服务的解决方案，包括车辆平台、硬件平台、软件平台、云端数据服务四大部分，具备环境感知、路径规划、车辆控制等功能以及强大的车载操作系统。

智能城市：传统的信息化建设导致我国很多部门积累的大量数据无法进行共享。阿里云公司开发的“城市大脑”（图 4-2）利用人工智能技术，使得很多数据可进行互通和协同，帮助城市进行交通态势评价与信号灯控制优化、城市事件感知与智能处理、公共出行与运营车辆调度，以及帮助进行社

小贴士

目前人工智能已经在工业、医疗、社交、交通以及金融等领域得到广泛应用，并深刻地影响着人们的生活。人工智能不仅能帮助人类完成很多重复枯燥和危险的工作，还能比人类做得更快、更准确，比如在特定任务中，语义识别、语音识别、人脸识别、图像识别技术的精度和效率已超越人工。

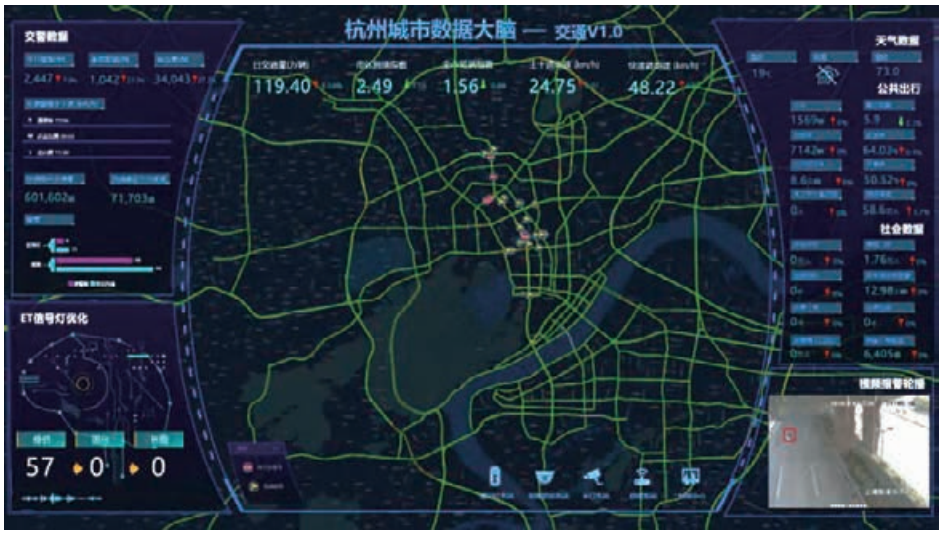


图 4-2　“城市大脑”应用案例

会治理与保障公共安全。“城市大脑”可以帮助城市管理者从全局角度实时发现城市的问题并给出优化处理方案，与城市内各项资源调度联动，从而提升城市的整体运行效率。

智能医疗：人工智能对医疗的推动作用主要体现在替代医生完成部分工作从而实现自动化医疗，辅助医生提高自身的工作效率，提高患者自查率，以利于更早发现、更准确地诊断疾病。有些医学影像智能筛查系统（图 4–3）能实现对早期食管癌、肺癌、乳腺癌、糖尿病性视网膜病变等疾病的智能化筛查和识别，辅助医疗临床诊断。

图 4–3　某医学影像系统

智能语音：语音是人类最自然、最重要的信息交流方式。得益于人工智能的发展，智能语音技术近年来取得了一系列突破性进展，并逐渐成为各智能设备的标配技术。智能语音能够实现人机间的语音通信，它包括语音识别技术和语音合成技术。有些智能语音平台具备语音合成、语音识别、口语评测（图 4–4）、语言翻译、声纹识别、自然语言处理等技术及能力。

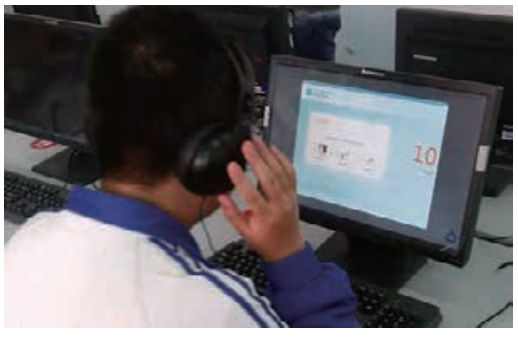


图 4–4　智能英语听说测试

未来，人工智能与各行业领域的深度融合将改变甚至重新塑造传统行业（图 4–5）。例如人工智能在制造、家居、金融、交通、安防、医疗、物流等行业的应用会使这些行业逐渐转变到智能制造、智能家居、智能金融、智能交通、智能安防、智能医疗和智能物流。

图 4-5　人工智能与传统行业的融合

虽然我们不一定时刻意识到人工智能的存在，但是人工智能已经深入到社会生活的方方面面。随着人工智能技术的不断发展，及其在各个领域应用的不断成熟，未来人工智能技术将进一步渗透到各行各业，对人类社会产生巨大的影响，并带来深刻的变革。

思考与讨论

你使用过哪些人工智能产品？人工智能给你的生活带来了怎样的改变？

活　动

9.1 百度、阿里云、腾讯和科大讯飞四家公司组成了我国首批人工智能的支柱企业，它们分别致力于自动驾驶、城市大脑、医疗影像和智能语音。以小组为单位，选择其中一个公司，收集它们在相应人工智能领域的最新产品，分析这些产品产生的影响，以及可能的发展前景。

2. 直面人工智能的安全、伦理问题

随着人工智能技术的不断成熟，越来越多的智能机器或智能程序成为人类的助手。人工智能系统作为一种信息系统，也可能面临一些安全风险。这些安全风险，既指人工智能系统自身的安全问题，如系统断电、部件损坏等，更指人工智能系统给人类社会带来的安全问题，及伦理问题。

（1）人工智能的系统安全风险问题

小贴士

伦理是指在处理人与人、人与社会的相互关系时应遵循的道理和准则，是指一系列指导行为的观念。

信息系统可能面临的安全风险包括：硬件风险、软件风险、数据风险、网络风险等。人工智能系统作为一种典型的信息系统，同样会面临上述的安全风险。

从人工智能系统本身来看：受限于人工智能技术的成熟度，某些技术缺陷有可能会引发相应的安全隐患。众所周知，人工智能系统依靠代码实现，这些代码是程序员写出来的，难免存在漏洞。一旦黑客利用漏洞进行攻击或劫持，人工智能系统可能被黑客控制，从而带来严重的后果。例如军事用途的人工智能系统如果有可能被不法分子劫持，导致战争的出现。此外，人工智能所依赖的传感器、训练数据和使用的开源软件等可能给人工智能系统带来安全风险，比如传感器可能被干扰，导致采集的数据出错。

从人工智能的应用方面来看：目前主流的基于机器学习的人工智能系统的一个重要特点是，系统的设计实现及运行无须进行过多的人类干预，能够实现自我学习调整。在这种模式下，人工智能系统的整个决策过程往往不需要操控者的指令，若其决策结果对人类社会产生危害，后果可能非常严重。

图 4-6　刷脸门禁

不仅如此，即使人工智能系统在功能上没有任何问题，仍可能存在应用方面的具体问题。有些人工智能系统不需要百分之百的准确率，如语音识别，99.99% 的准确率已经非常完美。而有些人工智能系统则需要 100% 的精确，比如智能刷脸门禁（图 4-6）系统绝对不能出现识别错误，既不能将房主拒之门外，也不能让非房主入门。

思考与讨论

你认为人工智能系统还存在哪些安全问题？参考信息系统的安全保护措施，试举例说明如何确保人工智能系统的安全。

（2）人工智能的伦理问题

人类区别于机器的一点是受伦理道德的规范和约束。随着人工智能技术的发展，机器的智能化程度越来越高，且逐步拥有了高层的感知、决策和反馈能力，此前只有人类才能完成的一些复杂认知任务正逐步交由机器完成。如何规范人工智能，让机器受人类伦理道德的约束是我们需要面对的挑战。

比如，自动驾驶中涉及这样的伦理问题：假设自动驾驶汽车正在高速行驶，前方突然出现行人，此时刹车失灵或者

无法立即停车，只能通过方向盘控制车辆，如图 4-7 所示，自动驾驶汽车面临两种选择：

A——任汽车继续行驶，这可能会伤害行人；

B——转弯冲向路边的障碍物，这可能会伤害自己。

在日常驾驶中，上述极端情况很少会出现，但是对于自动驾驶来说，必须在设计时就要考虑，一旦发生此类情况，要有符合道德规范的应对措施。面对未来，我们应设计一系列符合人类道德的准则来促进自动驾驶技术的发展。

图 4-7　撞人还是撞障碍物？

活　动

9.2 为了应对自动驾驶中可能出现的伦理问题，德国政府推出了一套自动驾驶道德伦理标准。请查阅该标准，思考由政府出台伦理标准的意义。

（3）人工智能的隐私问题

人工智能的应用中也会涉及对用户个人信息的合理使用问题。合理使用个人信息会让人工智能技术很好地为人类服务，然而非法使用个人信息则会造成对个人隐私的侵犯。

思考与讨论

如果所收集的大量个人隐私信息被非法使用，会对我们造成怎样的危害？

一些人工智能系统会收集用户的指纹、虹膜等生理特征来辨别人的身份，或收集用户的行为习惯来进行分析，之后给用户推送广告信息。有些网站利用人工智能技术，通过分析用户的浏览数据，归纳出用户的行为特征和习惯，进而给用户推送可能感兴趣的内容来提高点击量。这些应用虽然很“懂”用户的心思，但有关信息是在用户不知情的情况下收集的。

不仅如此，人工智能技术还有可能从已获取的个人信息中推导出用户不愿意泄露的隐私，例如从公共数据中推导出私人信息，从某个人的信息中推导出和这个人有关的其他人员（如朋友、亲人、同事）的信息（在线行为、人际关系等），这类信息往往超出了最初用户同意披露的个人信息范围。

小贴士

个人隐私信息的采集、分析、存储和使用都应得到个人的授权。

活　动

9.3 通过互联网搜索近年来发生的隐私泄露事件，总结泄露的隐私内容（如个人账号、个人信息等）、泄露的原因（如数据被盗、由于监管不善数据被违规使用等）、隐私被如何滥用（如出售、违规使用等）及其社会影响（如提高了大众的隐私保护意识、促进了更严的监管等）。

（4）保证人工智能的安全

为保证人工智能安全，我们可以采取一些基本的安全防护方法来确保人工智能系统安全。硬件方面，我们可以为人工智能系统安装物理防护，严格控制各种类型的物理访问，并设置备用设备，一旦原始人工智能系统遇到风险，可即刻启用备用设备。软件及网络方面，我们应该尽量减少软件可能存在的漏洞，一经发现应立刻进行修补，并通过权限管理、防火墙设置、通信加密等方式来降低网络风险。数据方面，我们应当及时做好数据备份，严格控制数据的授权及使用，并对数据进行加密等。

人工智能是人类智能的延伸。社会公众必须相信人工智能技术能给人类带来的利益远大于伤害，才有可能推动人工智能的发展。为了保障安全，人工智能技术本身及在各个领域的应用应遵循人类社会所认同的伦理原则。

比如人类利益原则：即人工智能应以实现人类利益为终极目标。这一原则体现对人权的尊重、对人类和自然环境利益最大化以及对社会危害性的重视。

又如责任原则：即在技术开发和应用两方面都建立明确的责任体系，以便在技术层面可以对人工智能技术开发人员或部门问责，在应用层面可以建立合理的责任和赔偿体系。在责任原则下，在技术开发方面应遵循透明度原则，即人类应该知道人工智能的工作原理；在技术应用方面则应当遵循权责一致原则，即政策和法律应对有关的权利和责任作出明确规定。

同时，我们应不断地探索设定人工智能的伦理准则，以确保人工智能在造福人类的同时也是安全可控的，不会威胁到人类本身的生存。

小贴士

建立令人工智能技术造福于社会、保护公众利益的政策、法律和标准化环境，是人工智能技术持续、健康发展的重要前提。

同时，作为开发者或用户，我们也应提升开发和使用人工智能系统的安全意识。

数字化学习

上网查找有关人工智能的伦理准则。

2017 年 12 月，电气电子工程师协会（IEEE）发布了第 2 版《人工智能设计的伦理准则》，为人工智能建立了社会与政策方面的指南，从而确立了人工智能系统能够以人为本，并服务于人类的价值和伦理准则。

思考与讨论

还可制定哪些共识原则来确保人工智能安全可控？

我国对于人工智能发展过程中的安全问题也给予了高度重视。2017 年 7 月国务院印发的《新一代人工智能发展规划》中明确指出，人工智能发展的不确定性带来新挑战，在大力发展人工智能的同时，必须高度重视可能产生的安全风险挑战，加强前瞻预防与约束引导，最大限度降低风险，确保人工智能安全、可靠、可控地发展。

活　动

9.4 分小组查阅目前关于人工智能安全的已经颁布的法律法规或行业标准。

3. 展望人工智能的未来

人工智能技术正在快速发展，并为世界带来全新变革。然而任何科学技术都是双刃剑，在给我们带来利好的同时也不免带来一些威胁。因此，需要辩证地看待人工智能的巨大价值和潜在威胁。

（1）人工智能的巨大价值

人工智能的系统应用已为人类创造出可观的经济效益。随着相关技术的不断发展与完善，人工智能技术必将得到更大的推广，产生更多的经济效益。近年来，人工智能几乎渗透到工业、医疗、社交、智能交通以及金融等各个领域，有力地促进了社会和经济的发展。

（2）人工智能的潜在威胁

根据人工智能的发展程度和实现的功能，人工智能的发展可以分为弱人工智能（Artificial Narrow Intelligence，ANI）、强人工智能（Artificial General Intelligence，AGI）和超人工智能（Artificial Super Intelligence，ASI）三个阶段。

参见 P121 知识链接“人工智能三阶段”

虽然我们都非常期待超人工智能的到来，但目前，还仅仅处于弱人工智能阶段。当前学者们对于超人工智能能否及何时到来尚无定论，对于人工智能是否会对人类产生威胁也有不同观点。持乐观看法的人认为人工智能在未来的很长一段时间内都将只是人类的一种工具；人类智能的门槛很高，人工智能很难在短时间内突破和超越人类智能。持悲观看法的人则认为人工智能技术不断加速发展的趋势已经无法改变，超越人类的超级智能将会在不远的将来出现，而那时人类将面临灾难性的生存危机。

思考与讨论

有人认为人工智能未来可能会威胁到人类的生存，提出了人工智能威胁论。你对人工智能威胁论是怎么认识的呢？

无论超级智能是否能到来及何时能到来，我们都需要对之进行充分研究，采取防范措施来确保人工智能会朝着安全可控且有利于人类的方向发展。

（3）人工智能的发展趋势

根据现有的发展情况，人工智能技术未来可能呈现出以下发展趋势。

① 更加通用的智能

目前主流的人工智能研究集中在弱人工智能，一般只擅长于单个方面的任务，比如 AlphaGo 只擅长于下围棋，而不擅长其他任务。但是人们对人工智能的期待不仅仅是解决特定的、单一的问题，而是像人类一样能够解决复杂的、不同领域的问题，能够进行推理和决策，即通用人工智能。尽管研究者对通用人工智能能否最终实现及是否真的安全可控尚无定论，但是人工智能由专用智能向通用智能的方向发展是未来的必然趋势，也是人工智能研究中极具挑战性的问题。

② 更小的数据及模型

现阶段人工智能 + 大数据的模式取得了极大的成功，海量的数据及强大的计算能力都极大地促进了人工智能技术的发展。在这种模式下，人工智能技术高度依赖于大量的标定数据，用以进行训练。但是数据标定过程费时费力，而且对有的问题难以采集到大量的数据，因此人工智能在未来的一个发展趋势是在少量数据甚至数据缺失的条件下进行学习。

另一方面，现阶段的人工智能模型太复杂，这使得相关技术无法在轻量级环境下得到大规模应用。因此，人工智能的另一个发展趋势是在保证系统性能的前提下，开发更小、更简单的模型，使得小模型能够达到与大模型相同的效果。

③ 更多的学科交叉

近年来，人工智能的研究反映了一个趋势：来自神经及脑科学的启发能够有效地提升现有人工智能模型的智能水平。然而，想要真正达到甚至超越人类的智能水平，还需要对脑科学进行更为深入的研究和借鉴。此外，如前文所述，人工智能技术在给我们带来极大便利的同时，也对人类现有的法律、道德及伦理体系带来极大的冲击。人工智能不再仅仅是一个单纯的技术问题，还需要社会学、伦理学等学科的研究人员对相关问题进行深入的研究。

活　动

9.5 查阅相关资料，总结人工智能的发展趋势。选择一个方向，畅想未来可能出现的新技术、新应用、新变化等。

9.6 从以下题目中选择一个，或自拟一个题目，组织一场辩论会。

参考题目：

- 人工智能会再次进入冰冻期吗？
- 人工智能真的智能吗？
- 人工智能会超越人类智能吗？
- 人工智能是隐私的天敌吗？

知识链接

人工智能三阶段

1. 弱人工智能

又称狭义人工智能，是指能推理和解决问题的智能机器。这种机器不会有自主意识，一般只擅长单方面的任务。

2. 强人工智能

又称广义人工智能，是指真正能推理和解决问题的智能机器。这种机器具有知觉和自我意识，它是跟人类智能级别接近的人工智能，在各方面都能和人类比肩。

3. 超人工智能

一般简称超级智能，在几乎所有领域都比人类强。理论上，强人工智能发展成为超人工智能比较简单，因为只要将强人工智能的认知能力再强化一点点，它就可能在各方面都超越人类。

拓展阅读

人工智能伦理准则

IEEE“人工智能设计的伦理准则”包含宗旨、目标、目的、理论基础、未来技术关切五大部分：

- 其宗旨是建立框架，指导我们认识这些技术可能造成的技术以外的影响，并就此进行对话和讨论。
- 其目标是在遵循人权、福祉、问责、透明、慎用的一般原则下合乎伦理地设计、开发和应用人工智能技术。
- 其目的包括个人数据权利和个人访问控制、通过经济效应增进福祉、问责的法律框架、透明性和个人权利、教育和知悉的政策五个方面。
- 其理论基础包括经典伦理学、福祉指标、将价值嵌入自主系统、指导合乎伦理的研究和设计的方法四个方面。
- 其未来技术关切包括重新定义自主武器、所谓的通用人工智能和超人工智能的安全性和有益性、情感计算、混合现实四个方面。

——摘自 IEEE“人工智能设计的伦理准则”

人工智能发展原则

2017 年 1 月，在加利福尼亚州阿西洛马举行的 Beneficial AI 会议讨论了人工智能的未来及其监管。近千名人工智能专家联合签署了阿西洛马人工智能 23 条原则，并号召人工智能的研究人员、科学家和立法者遵循这些原则，以确保人工智能安全和有利于人类。

阿西洛马人工智能 23 条原则从方法、伦理、道德等方面限定了人工智能可以干什么不可以干什么。阿西洛马人工智能 23 条原则涵盖了三个范畴：

第一类为科研问题，共 5 条，包括研究目标、研究经费、科学与政策的联系、科研文化、避免竞争；

第二类为伦理和价值，共 13 条，包括安全性、故障透明性、司法透明性、责任、价值归属、人类价值观、个人隐私、自由和隐私、分享利益、共同繁荣、人类控制、非颠覆、避免人工智能军备竞赛；

第三类为更长期的问题，共 5 条，包括能力警惕、重要性、风险、递归自我完善、公共利益。

——摘自“阿西洛马人工智能原则”

单元挑战　设计无人驾驶时代的交通准则

一、项目任务

在无人驾驶时代，我们乘坐的无人驾驶车具有完全的自主性。无人驾驶车能够完成所有的驾驶操作，包括路径规划、自动驾驶、行人及车辆避让等。我们需要设计一系列的准则来指导自动驾驶系统的设计，并且让自动驾驶系统在出现任何情况时都能够有章可循，保障人工智能时代的道路安全。

二、项目指引

1. 收集资料，了解无人驾驶时代应该从哪些方面设计交通准则来保障道路交通安全，例如技术、法律、伦理等方面。

2. 确定设计的方面，分别从这几个方面进行深入思考和挖掘，并设计若干条相应的准则，并加以梳理和归纳。

方面	准则

三、交流评价与展示

1. 以思维导图的形式展示设计的交通准则。

2. 征询大家对每条准则的意见，看每条准则是否能够获得普遍认可。

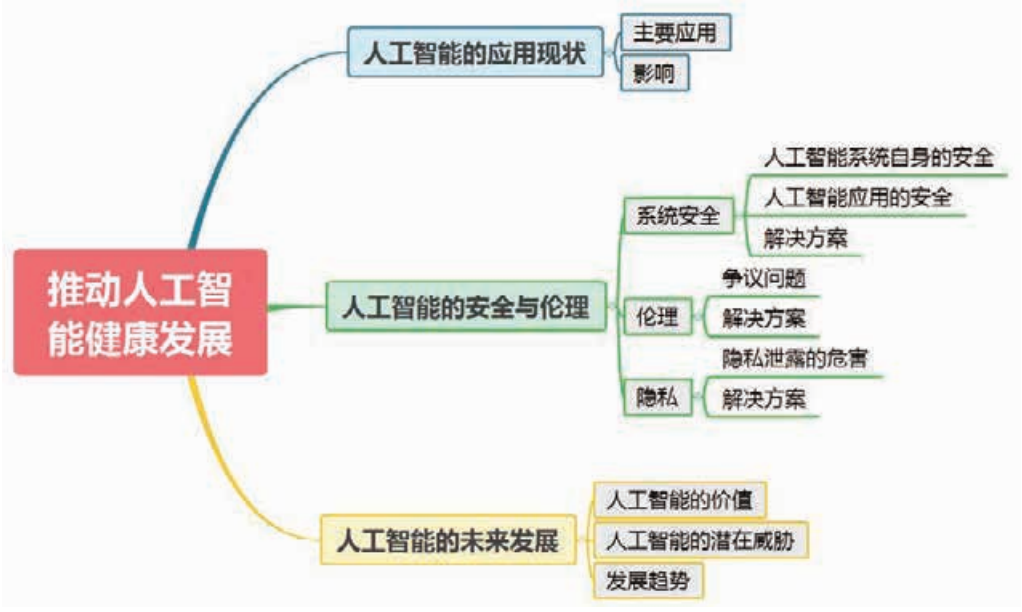
3. 交流并学习他人设计的交通准则（尤其是自己没有考虑到的方面），总结大家设计准则的共性。

设计的共性：

4. 回顾无人驾驶交通准则的设计过程，反思并交流期间的收获与不足。

单元小结

一、主要内容梳理



二、单元评价

评价内容	达成情况
能列举人工智能技术在日常生活中的应用（A、R）	
能从多个角度描述人工智能对人类社会的影响（A、R）	
能说出智能驾驶、智能城市、智能医疗和智能语音对我国的意义（A、R）	
能说出我国首批人工智能平台的功能和意义（A、R）	
能说出人工智能系统的常见系统安全问题（A、R）	
能举例说明人工智能中涉及的伦理道德（A、R）	
能说出确保人工智能安全的技术方法和伦理准则（A、R）	
能用自己的话阐述人工智能的价值和潜在威胁（A、R）	
能用自己的话阐述人工智能的发展趋势（A、R）	

说明：A—信息意识，T—计算思维，I—数字化学习与创新，R—信息社会责任

附 录

部分名词术语中英文对照

（以汉字拼音字母次序为序）

残差网络 ResNet
长短期记忆网络 Long Short Term Memory，LSTM
超人工智能 Artificial Super Intelligence，ASI
对角线距离 diagonal distance
代价函数 cost function
反向传播 Back Propagation, BP
非叶节点 nonleaf node
非线性回归 nonlinear regression
分类 classification
感受野 receptive field
根节点 root node
广度优先搜索 breadth-first search
过拟合 over-fitting
机器学习 Machine Learning, ML
激活函数 activation function
计算能力 computing power
监督学习 supervised learning
局部最优解 local optima
聚类算法 clustering algorithm
聚类中心 cluster center
卷积神经网络 Convolutional Neural Networks,CNN
决策树 decision tree
kNN 算法 k-Nearest Neighbor algorithm
k- 均值聚类算法 k-means clustering algorithm
路径规划 path planning
LBP 算法 Local Binary Pattern algorithm
曼哈顿距离 manhattan distance
梅尔倒谱系数 Mel-Frequency Cepstral Coefficients,MFCC
欧氏距离 euclidean distance
PCA 算法 Principal Component Analysis algorithm
启发式搜索 heuristic search
欠拟合 under-fitting
强化学习 reinforcement learning
强人工智能 Artificial General Intelligence，AGI
人工智能 Artificial Intelligence, AI
人脸识别 face recognition
弱人工智能 Artificial Narrow Intelligence，ANI
深度学习 Deep Learning,DL
深度优先搜索 depth-first search
神经网络 Neural Network, NN
神经元 neuron
树莓派 Raspberry Pi
特征 feature
特征向量 feature vector
梯度下降 gradient descent
图灵测试 The Turing Test
图像识别 image recognition
图形处理单元 Graphics Processing Unit,GPU

微控制单元　Micro Control Unit,MCU

文本转语音生成　Text To Speech, TTS

无监督学习　unsupervised learning

现场可编程逻辑门阵列　Field Programmable Gate Array,FPGA

线性回归　linear regression

循环神经网络　Recurrent Neural Network,RNN

叶节点　leaf node

隐马尔可夫模型　Hidden Markov Model, HMM

语音合成技术　Speech Synthetic Technology, SST

语音识别　Automatic Speech Recognition, ASR

支持向量机　Support Vector Machine, SVM

中央处理器　Central Processing Unit,CPU

专家系统　Expert System, ES

专用集成电路　Application Specific Integrated Circuit,ASIC

子节点　child node

自回归滑动平均　Auto Regressive Moving Average，ARMA

自然语言理解　Natural Language Understanding, NLU

自然语言生成　Natural Language Generation, NLG

最小二乘法　least-squares method

PUTONG GAOZHONG JIAOKESHU
XINXIJISHU

普通高中教科书
信息技术　选择性必修4
人工智能初步
上海科技教育出版社有限公司出版发行
（上海市闵行区号景路159弄A座8楼　邮政编码201101）
湖南省新华书店经销　湖南长沙鸿发印务实业有限公司印刷
开本890×1240　1/16　印张8.25
2021年1月第1版　2021年12月第3次印刷
ISBN 978-7-5428-7413-9/G·4351
定价：10.37元
批准文号：湘发改价费〔2017〕343号　举报电话：12315

此书如有印、装质量问题，请向印厂调换
印厂地址：长沙黄花印刷工业园三号　电话：0731-82755298