普通高中教科书

信息技术

选择性必修4

人工智能初步

浙江教育出版社
ZHEJIANG EDUCATION PUBLISHING HOUSE

普通高中教科书

信息技术

选择性必修4

人工智能初步

闫寒冰 主编

浙江教育出版社·杭州

主　　编：闫寒冰
副 主 编：赵　健　魏雄鹰

本册主编：吴　飞

前　言

信息技术作为当今先进生产力的代表，已经成为我国经济发展的重要支柱和建设网络强国的战略支撑。在这样的大背景下，教育部全面修订并颁布了《普通高中信息技术课程标准（2017年版）》，为这门课程设定了与新时代相符的育人目标：帮助学生掌握信息技术基础知识与技能、增强信息意识、发展计算思维、提高数字化学习与创新能力、树立正确的信息社会价值观。

本套教材依据《普通高中信息技术课程标准（2017年版）》编写，包括两本必修教材《数据与计算》《信息系统与社会》，六本选择性必修教材《数据与数据结构》《网络基础》《数据管理与分析》《人工智能初步》《三维设计与创意》《开源硬件项目设计》，两本选修教材《算法初步》《移动应用设计》。

本套教材的编写组汇集了来自信息技术、课程与教学、教育技术等领域的高校学者与教学一线专家。编者们通力合作，从课程内容、教材体例、技术选择、教学方法、学习方法等方面精心打磨，期待以最专业的样态帮助学生达到课程预期的育人目标。

具体而言，本套教材体现了如下特点：

1. 体例上——为核心素养的培养创造空间和条件：将核心学习内容与支持学习的方法有机融合在一起，支持学生在自主、合作、探究的学习情境下发展核心素养。

2. 内容上——体现概念、内容与方法的精准与专业：在增强教材可读性的同时，精炼提升综合素养所必需的核心内容，强调所有概念、内容与方法的精准与专业。

3. 活动上——着力提升学生的高级思维能力：精心设计与布局教材中的练习、思考、讨论、实践与项目学习，追求对高级思维能力的培养。

4. 案例上——体现信息科技的多层需求与多维格局：把案例的呈现作为开阔视野的重要手段，帮助学生理解信息技术对于社会发展所具有的价值与意义。

5. 技术上——引领学生拓宽视野与发展思维：将每种具体应用软件都作为解决某些问题的一条路径来看待，期待学生通过具体的技术操作体验，理解其背后的原理与格局、特点与局限，拓宽视野、发展思维。

本册教材为选择性必修《人工智能初步》。通过本教材的学习，期待同学们能了解人工智能的历史发展、基本算法和计算模型，知道人工智能在社会生活中所起的重要作用，理解人工智能伦理与安全，运用人工智能解决生活与学习中的特定问题和特定任务，认识到人工智能在信息社会中越来越重要的作用，增强利用智能技术服务人类发展的责任感。

就教材本身所讲述的知识内容而言，我们相信，只要同学们潜心自学就可以基本掌握。但“知识内容”只是发展信息技术核心素养的基础部分，所以，我们希望同学们不要仅满足于对具体知识与具体技术的掌握，还要重视教材中的各类学习活动，与老师和学友一起，更多地去创造、研究、解决问题、制作、交流、合作和评价，唯有如此，同学们才能藉由这门课程的学习全面地提升信息素养，增强在信息社会的适应力与创造力，为实现中华民族伟大复兴的宏伟目标做出更大贡献！

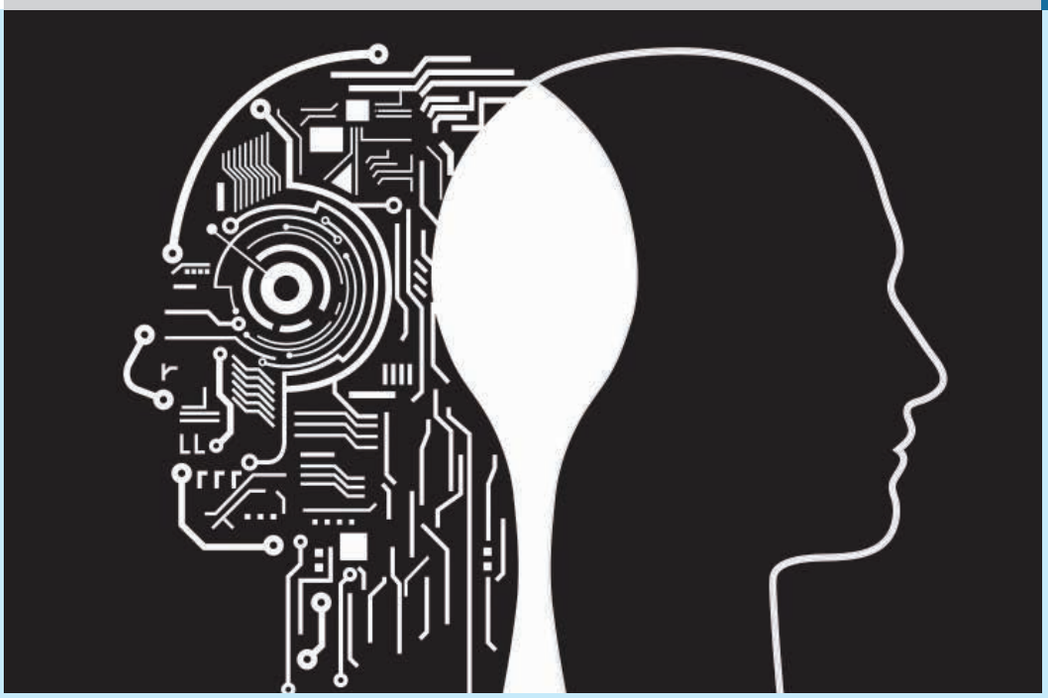
本册教材在编写过程中得到了各方面的大力支持。浙江大学计算机学院钱沄涛教授、李玺教授、王志坚研究员、钱徽教授、金小刚副教授、汤斯亮副教授、赵洲副教授和杨洋副教授为本书的撰写提供了建议。浙江大学计算机学院焦云皓、张圣宇、周乐夔、廖彬兵、林宇箫、宋骏、蒋胤傑、段新宇、李蕴哲等研究生参与了编撰工作。来自浙江湖州的韦黎沁、孙杰、傅海涛和李宝华四位高中教师对教材的部分内容提出了修改意见。西安交通大学郑南宁院士、上海交通大学施鹏飞教授、南京航空航天大学陈松灿教授、浙江大学计算机辅助设计与图形学国家重点实验室金小刚教授、西安交通大学兰旭光教授、天津大学韩亚洪教授、兰州大学马志新教授、武汉大学肖春霞教授、复旦大学金城教授、西北工业大学聂飞平教授在百忙之中对书稿内容进行了审阅。

由于水平有限，本书可能还存在不足之处。希望大家在教材使用过程中，能够及时将意见和建议反馈给我们，对此，我们深表谢意。

目　录
MULU

第一章 1

智能之路：历史与发展



人工智能（Artificial Intelligence，简称AI）是以机器为载体所展示的人类智能，因此人工智能也被称为机器智能（Machine Intelligence）。

人类一直不懈努力，让机器模拟人类智能，从而提升生产能力、帮助人类完成更为复杂或有危险的工作，更多造福人类社会。迄今为止，对人类智能的模拟存在形式化推导和数据驱动学习两条主要实践道路。形式化推导主要是通过逻辑推理的方法从若干判断（前提）出发得到新判断（结论），逻辑推理为以规则为核心的人工智能发展奠定了理论基础。数据驱动学习则从数据出发，发现数据中所蕴含的模式和规律。前者可视为“从知识到知识”，后者可视为“从数据到知识”。

无论是形式化推导，还是数据驱动学习，均是“计算”过程，需要通过具体计算平台实现。20世纪中叶，学术界对“可计算问题”产生了浓厚兴趣，在研究什么问题可以被计算、什么问题不可以被计算的过程中，阿兰·图灵（Alan M. Turing）提出了图灵机模型。这个模型成为现代计算机的理论模型，也成为实现当今人工智能的机器载体。

图灵曾经提出了“图灵测试模型”来判断人类智能和机器智能，当前人工智能方法无法让机器产生自我智能或意识。人工智能在发展过程中逐渐形成了其内涵，包括问题求解、知识表达、语言理解、视觉计算、机器学习与机器协调控制等内容。与单纯机器智能不同，当前科学家也提出了脑启发智能、混合增强智能等智能形态，表现为外骨骼机器人、“人—机—物”融合三元空间和智慧城市等应用。

问题与挑战

- 人工智能的发展目标是模拟和仿真人脑所具有的智能活动。在发展人工智能过程中，存在如下三种思路：第一种主张要在完成解析人脑全部机理后，才能实现人工智能；第二种思路是不用等待人脑机理被全部解析，而是走出一条不依赖于人脑机理的人工智能发展之路，如1901年莱特兄弟实现人类第一次动力飞行时，人类尚未建立空气动力学；第三种思路是将神经科学进展和计算机科学进展相互结合、共同发展。三种思路的内涵与本质分别是什么？

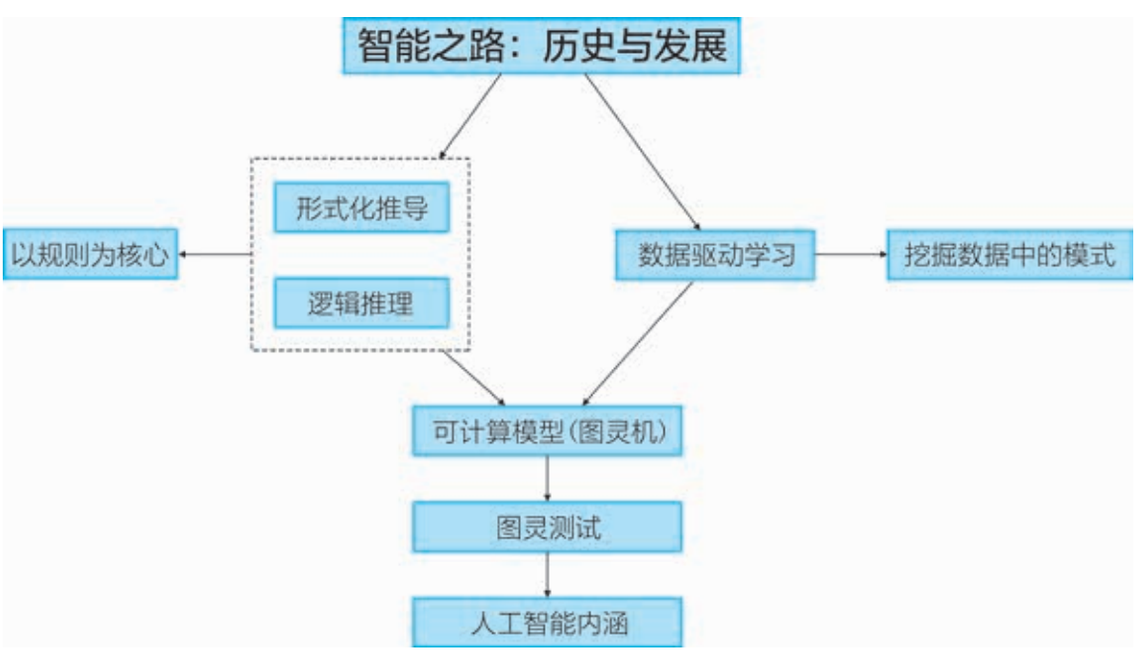
- “小任务、大数据”（计算机完成任务模式）和“大任务、小数据”（人脑完成任务模式）形象地描述了计算机和人脑在完成特定任务过程中存在的区别。比如，人类个体只要看几张飞机图像就可识别飞机，而计算机则需要从成千上万张飞机图像中进行识别学习。计算机和人脑在完成智能任务时为什么会采用上述不同的方法？

- 蝙蝠在黑夜通过超声波回声定位信号来避让前方障碍物和逃避敌害等。人类用电磁波替换了超声波来探测障碍物，从而发明了雷达，可以说是借助生物智能更好地实现了机器智能。生活中还有哪些借助生物智能实现的机器智能？

学习目标

1. 了解人工智能的概念和内涵。
2. 掌握逻辑推理、可计算和图灵机等基本概念。
3. 了解人工智能的发展历史、典型应用与趋势和局限性。
4. 了解人工智能开源平台基本情况，掌握Keras配置方法。

内容总览



1.1 人工智能的起源

1.1.1 古代智能思想探究

人工智能是以机器为载体所实现的生物智能（如鸟类或鱼类等所具备的能力）或人类智能。人是生物分类学上的智人（Homo sapiens，来源于拉丁语，意思是“wise man”，即“有智慧的人”），是一种灵长目人科人属的直立行走物种。一般而言，人类智能具有非常广泛的内涵，如联想、顿悟、判断、决策等。

古今中外众多思想家从思辨和抽象的角度分别思考过智能的内涵，并给出了精辟论述。《道德经》中写道：“知人者智，自知者明。胜人者有力，自胜者强。知足者富，强行者有志。”老子认为能了解他人的人是聪明的。

《论语·子罕》中写道：“知（注：“知”通“智”）者不惑，仁者不忧，勇者不惧。”孔子将智慧、仁义和勇敢列为传统道德中三个重要范畴，认为有智慧的人不会迷惑。

《孙子兵法》中将“智”列在“将”的五德之首，认为“智能发谋、信能赏罚、仁能附众、勇能果断、严能立威，五德皆备，然后可以为大将”。这反映了孙子在军事对决中强调“智将”而非“勇将”，即不战而屈人之兵是首选，而不是攻城拔寨的“勇夫”。

孟子在“四端”中对“智”进行了如下阐述：“恻隐之心，仁之端也；羞恶之心，义之端也；辞让之心，礼之端也；是非之心，智之端也。”

《荀子·正名》中给出了“智能”的定义：“所以知之在人者谓之知。知有所合谓之智。智所以能之在人者谓之能。能有所合谓之能。”这句话揭示了从感知到理解，然后到认知，最后到决策与行动的智能链条。

柏拉图在《理想国》中把“正义、智慧、勇敢和节制”并列为人类的四大美德。

可以看出，古今中外的哲人都从道德、心智、认知等角度来讨论智能，但与智能能力模拟还相去甚远。

1.1.2 逻辑与推理

推理是进行思维模拟的基本形式之一，是从一个或几个已知的判断（前提）推出新判断（结论）的过程。因此，只要前提正确，推理所得出的结论往往是正确的。

亚里士多德提出和建立的“演绎三段论”（syllogisms）是一种著名的逻辑推理手段。简单地说，演绎三段论从大前提和小前提出发，推理出结论。大前提是一般性原则，小前提是一个特殊陈述。下面给出两个演绎三段论的例子。

●●● 例1

所有的恒星都是气态星体（大前提）
太阳是恒星（小前提）
太阳是气态星体（结论）

●●● 例2

所有心学学派人士都主张“知行合一”（大前提）
王守仁是心学学派人士（小前提）
王守仁主张“知行合一”（结论）

三段论推理只能出现三个概念，如例1中“恒星”“气态星体”和“太阳”是三个概念，大前提和小前提中共同出现了“恒星”这一概念。

在三段论演绎推理中，为了保证推理结果的正确性，需要保证大前提和小前提的正确性。即三段论的原则是大前提正确，小前提正确，则结论是正确的。下面给出一个通过演绎三段论得出错误结论的例子。

●●● 例3

所有的鸟都会飞（错误大前提）
鸵鸟是鸟（小前提）
鸵鸟会飞（错误结论）

例3中“所有的鸟都会飞”这一大前提是错误的，因此导致推理所得结论也是错误的。

下面给出另外一个通过演绎三段论得出错误结论的例子。

●●● 例4

中国的大学是分布于全国各地的（大前提）
浙江大学是中国的大学（小前提）
浙江大学是分布于全国各地的（错误结论）

例4依据三段论推理得出的结论是错误的。其原因在于大前提中的“中国的大学”是集合概念，小前提中的“中国的大学”是非集合概念。因此，在这个三段论演绎推理中，出现了四个不同的概念，导致推理错误。

中国古代也出现了形式逻辑的雏形，如《墨辩·经说上》中有“小故，有之不必然，无之必不然”“大故，有之必然，无之必不然”。这里的“小故”和“必要条件”非常相似（“小故”就是“有之不必然，无之必不然”），“大故”和“充要条件”非常相似（“大故”就是“有之必然，无之必不然”）。

人工智能研究与人的思维研究紧密相关，因此对逻辑推理的研究有力促进了人工智能

发展。亚里士多德有关逻辑的论著被后人结集为《工具论》，形成了包含归纳、演绎等方法的较为独立和系统的逻辑推理理论，因此亚里士多德被誉为“逻辑学之父”。

思考与练习

为使下述推理成立，必须补充以下哪项作为前提？写出三段论推理。

有些导演留大胡子，因此，有些留大胡子的导演是大嗓门。

A. 有些导演是大嗓门　　B. 所有大嗓门的人都是导演

C. 所有导演都是大嗓门　　D. 有些大嗓门的不是导演

1.1.3　可计算思想产生

《庄子·天下》中论述道：“一尺之捶（“捶”通“棰”），日取其半，万世不竭。”虽然从极限角度而言，该过程的结果是零，但是这个过程在哲学意义上是永不停息的，是有限和无限的统一。

人工智能赋能实体经济的能力依赖于机器载体，即计算机本身。任何事物的产生均有理论准备，计算机这一工具的诞生来源于对可计算思想的研究。

1900年，数学家希尔伯特在第二届国际数学家大会上提出了其认为具有重要意义的23个数学问题，其中被称为“判定丢番图方程的可解性”的“希尔伯特第十个问题”对推动计算思想的发展产生了深远的影响。

“希尔伯特第十个问题”可简单表述如下：是否存在一种算法，能够判定某个方程有解（即该问题是不是可计算的）？如是否存在一种算法可判定方程$x^n+y^n=z^n$（n为大于或等于3的正整数）存在整数解。

当时人们对可计算的本质是什么尚不清楚，难以精确定义什么问题是可计算的、什么问题是不可计算的。1937年，图灵发表了论文《论数字计算在决断难题中的应用》，在文中提出了被后人称为图灵机的计算机理论模型。因此，图灵被誉为“理论计算机之父”。计算机界的最高奖项图灵奖就是用其名字命名的。

拓展链接

图灵奖

图灵奖由美国计算机协会（ACM）于1966年设立，奖励对计算机事业做出重要贡献的个人，用计算机理论模型提出者阿兰·图灵命名。一般每年只奖励一位计算机科学家，只有极少数年度有两位合作者或在同一方向做出贡献的科学家共享此奖。它是计算机界最负盛名、最崇高的一个奖项，有“计算机界的诺贝尔奖”之称。

清华大学的姚期智先生因为在计算理论方向做出的杰出贡献（包括伪随机数生成、密码学与通信复杂度等）于2000年获得图灵奖。

图灵机模型是一个抽象的机械式计算装置，它有一条两端无限长的纸带，纸带被分隔成一个个可擦写的小方格，小方格中可放入数据、程序指令或者为空。在图灵机模型中，一个控制器控制一个机器头对纸带方格中的信息进行操作。在每一时刻，机器头从当前所指向纸带方格中读入该方格中的信息，然后根据该读入信息从控制器内部程序表中查找应对所读入信息的合理操作指示，根据该指示进行计算，并将计算所得的信息输出到纸带方格中，并转换内部状态，进行下一步计算。上述步骤周而复始，直至计算结束。一旦计算结束（即图灵停机），纸带上所记载的信息就是计算结果。

图1.1.1　图灵机模型计算“12+8”的过程示意图

图1.1.1给出了图灵机模型对“12+8”这一问题进行计算的过程。如图1.1.1甲所示，读入方格中“12”这个数字，指令集指示控制指针右移1个方格；如图1.1.1乙所示，读入加法符号，指令集指示控制指针右移1个方格，寻找被加数；如图1.1.1丙、丁所示，读入方格中“8”这个数字，指令集指示执行加法操作，并将相加结果“20”写入方格。可以看出，当“12+8”这一问题计算完毕后，只有原来数字“8”的方格上存在“20”这个数字，其他数字或程序信息均被擦除。

因为图灵机在若干步骤后停机，因此“12+8”是图灵可停机问题（即可计算的）。

在图灵机基础上，“可计算”概念可做如下定义：一个问题若能被图灵机经过有限步骤处理而停机，则该问题是可计算的。因此，在图灵机模型这个语境下，所有可计算问题都是图灵可停机问题、所有不可计算问题都是图灵不可停机问题。也就是说，若给定一个问题，图灵机在执行了有限多步骤后可停机输出计算结果，则这个问题是可计算的；否则，这个问题是不可计算的。

20世纪30年代，除了图灵机模型，还出现了原始递归函数（recursive function）和Lambda演算（Lambda calculus）两种计算模型。图灵机、原始递归函数和Lambda演算三者功能是等效的。也就是说，若一个问题能被图灵机模型计算，则该问题也能够被原始递归函数和Lambda演算计算。但是，与其他两个模型都要通过数学模型来实现不同，图灵机通过机械方式就能进行计算，因此成为现代计算机的理论基础。

拓展链接

原始递归函数与Lambda演算

原始递归函数是通过组合和原始递归运算所得到的一类函数。Lambda演算是用于研究函数定义、应用和递归的数学逻辑系统。邱奇—图灵论题（the Church-Turing thesis）指出：所有可计算问题都可以通过图灵机模型来解决。由于原始递归函数、Lambda演算与图灵机功能等效，因此所有可计算问题也可以通过原始递归函数和Lambda演算来计算。

思考与练习

1. 孪生素数是指取值相差为2的素数对，如3和5、17和19。华裔数学家张益唐证明了存在无穷多个差小于7000万的素数对。也就是说，如果两个素数之间的差值小于7000万，那么这样的素数对有无穷多。问：求解最大素数对（即该素数对中包含最大素数对）这个问题是可计算的吗？

2. 写出“3乘以5”这个问题在图灵机纸带上的计算过程。

3. 3月14日为圆周率日，是为了纪念南北朝时期杰出的数学家、天文学家祖冲之。他在魏晋时期数学家刘徽提出的“割圆术”思想的基础上，首次将“圆周率”精算到小数点后第七位。刘徽的“割圆术”思想是“割之弥细，所失弥少。割之又割，以至于不可割，则与圆周合体而无所失矣”。请用“割圆术”思想编写程序计算圆周率，体会可计算和逼近等思想。

4. 人们常说“万物皆可算”。现实生活中有很多可计算的例子，如100以内所有素数之和，从一张图像中识别出所有人脸等。但是，生活中也存在许多不可计算的例子，比如大数的质因子分解（把一个很大的数分解为质数相乘）等。请列举一些无法计算所得的例子。

1.2 人工智能的现状与发展

1.2.1 人工智能登上历史舞台

1955年，约翰·麦卡锡（1971年图灵奖获得者）、马文·明斯基（1969年图灵奖获得者）、克劳德·香农（信息理论之父）、罗切斯特（第一代通用计算机701主设计师）联名向美国洛克菲勒私人基金会提交了一份“人工智能”项目建议书。这份建议书中首次出现了“Artificial Intelligence”这个单词，明确提出了人工智能的研究目标：让机器能像人那样认知、思考和学习，即用计算机模拟人的智能。1956年，“人工智能”的概念在达特茅斯会议上引起了计算机学界的关注，人工智能开始走上人类历史发展的大舞台。

在“人工智能”项目建议书中，四位学者列举了人工智能面临的七个难题：（1）自动计算机；（2）计算机程序编程；（3）神经网络；（4）计算复杂性；（5）自我学习与提高；（6）计算的抽象能力；（7）随机性和创造力。

人类具有不断从环境中进行自我学习与提高的能力，具有归纳抽象和演绎推导能力以及直觉顿悟等创造性思维能力。这些能力是目前人工智能所欠缺的，成为迈向通用人工智能所面临的挑战性难题。

从20世纪70年代以来，科学家在机器定理证明、机器翻译、人机博弈、识图辨物等领域进行了研究。但是，人工智能的发展曾经历了三次大低谷，并非一帆风顺。

第一次低谷的发生缘于1973年英国数学家詹姆士·莱特希尔发表的报告。该报告对当时英国人工智能研究进行了评估，认为当时英国的人工智能研究集中在自动机、机器人和神经系统，并得出如下结论：自动机和中央神经系统的研究有价值，但进展令人失望；机器人的研究没有价值，进展非常令人失望，建议取消机器人的研究。遭此打击之后，人工智能进入严冬（AI Winter）。

第二次低谷由日本智能（第五代）计算机的研制失败导致。日本通产省1982年开始了第五代计算机的研制计划，希望计算机具备能直接推理与知识处理的新型结构。该计划的目标是构成一个具有1000个处理单元的并行推理机，连接10亿信息组的数据和知识库，具备听说能力。到1992年，该计划耗资约8.5亿美元，因没能突破关键性的技术难题，最终以失败告终。

第三次低谷始于1984年，当时的人工智能专家试图以专家手工构建方式，来生成包含人类所有知识的“知识大脑”，并期望基于此来对所有问题进行推理。当时人工智能专家认为，如果拥有了包罗万象、囊括万物的“知识大脑”，就可以通过“知识大脑”构建专家系统，回答任何问题。后来发现，将人类所有知识完整收集且形式化描述的任务根本无法完成。因此，单纯依赖规则和知识的人工智能系统（即专家系统）未能表现出所期望的效果。

1.2.2 智能测试

通过对算法进行测试，来判断其是否具有人类的智能水平是一个重要问题。图灵提出了一种称为“图灵测试”（the Turing test）的方案。

在图灵测试中，一台装载了智能算法的机器与一个人被分别安排在两个不同的房间。通过一些装置（如键盘）分别向机器和人随意提问，然后分别收集机器和人的回答结果。经过多轮提问后，将收集到的回答结果交给一个法官评判，请法官指出哪些回答是机器做出的、哪些回答是人做出的。如果法官无法有效区分机器和人的回答结果，那么可以认为这台机器具有智能。

从严格意义上说，目前尚未有任何智能算法或系统通过了“图灵测试”。一般认为，计算机完成一项特定任务（如刷脸支付或语音识别等）与计算机具有智能是两个不同的概念。人类智能具有极其广泛的外延，很难通过若干次问答结果正确与否来判断一个算法是否与人类智能相同。

图灵测试这个思想目前在生活中得到了应用，如在登录互联网系统时，系统经常会要求用户输入验证码，以此来判断是机器算法还是人类登录该系统。

“中文房间”实验是一个用来证明即使某个算法通过图灵测试，也难以评判其是否具有智能的实验。1980年，加州大学伯克利分校的某位哲学教授发表了一篇论文《意识、大脑与编程实现》，提出了一个“中文房间”实验，认为一个算法即使通过了图灵测试，也不意味着其具有智能，至多是对智能某个侧面的模拟。在这个实验中，一个从小讲英语但完全不懂中文的男士待在一个房间中。房间中有一本《汉英大词典》，假设这本大词典已经囊括了所有问题的答案。一旦房间外有人向房间内的男士递送中文书写的问题纸条，男士可以根据大词典以及若干规则找到中文纸条问题所对应的答案，进而输出中文答案，以至于房间外面的人认为房间内该男士熟知中文。

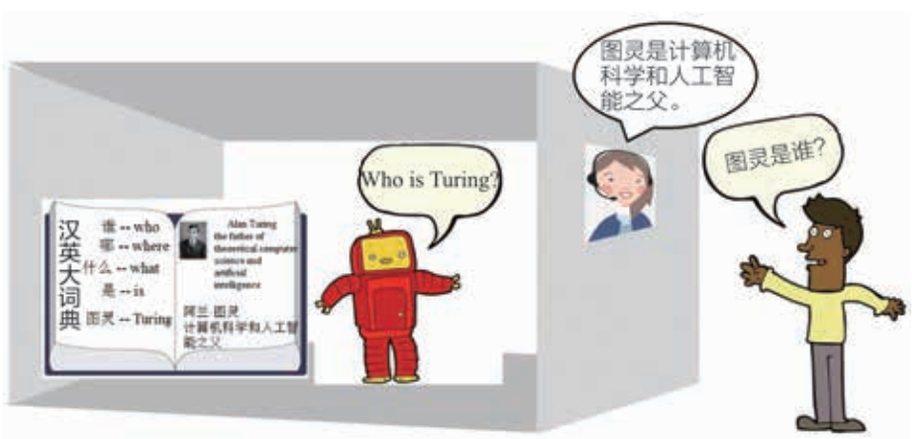


图1.2.1 问答测试的“中文房间”实验示意图

在上述中文房间实验中（如图1.2.1），房间内的男士（甚至是机器）都不必理解输入和输出字符的含义，只是按照一定规则机械地匹配字符符号，就给房间外的人留下理解中

文的印象，而实际上房间内的男士在问答过程中谈不上思维和理解。

举一个例子来说明通过词典机械式匹配存在的不足。图1.2.2给出了“德”这个汉字的一种演化形式（篆书）。假设房间内男士所携带的词典中没有“德”这个字及其含义的解释，那么其就无法输出“德”的含义了。但是，大家知道，“德”字各个部分有如下含义：十只眼睛、十只眼睛看到了直行、心也是直的、去观测人家在四通八达的大道是否直行。

图1.2.2　辨图中的推理

于是，从这些不同的构成元素，人类通过推理可以知道“德”的含义是道德或品行。人类思维不只是机械式的匹配过程，而且有丰富的推理和感悟。

问题与讨论

篆文“学习”两字如图1.2.3所示，按字面意思可知，在屋子里学习理论知识为“学”，在屋外从实践中学习为“习”。请发挥想象，以辨图推理的思想来分别理解“学”和“习”，然后从人工智能算法要不断从已有知识和实践锻炼中增长能力的角度来理解“学习”这一词语的含义。

图1.2.3　篆文“学习”

1.2.3　人工智能研究内容

人工智能具有强大的渗透力和支撑性，可用“至小有内，至大无外”来形容人工智能涉及的方方面面，即人工智能研究既有其核心内容，也有一些支撑内容，如智能教育、智能医疗、智能司法和智能金融等。

从模拟人类智能的角度而言，人工智能应该具备如下基本能力：

（1）视觉感知和语言交流的能力。即能够识别和理解外界信息（计算机视觉研究范畴）、能够与人通过语言交流（自然语言理解研究范畴）。

（2）推理与问题求解的能力。即基于已有知识，对所见事物和现象进行演绎推理以解决问题。

（3）协同控制的能力。即将视觉（看）、语言（说）、推理（悟）等能力统一协调，加以控制，这是常见的机器人研究领域内容。

（4）遵守伦理道德的能力。即模拟人类智能的智能体在社会环境中要遵从一定的伦理道德。

（5）从数据中进行归纳总结的能力。即需要从数据中获取知识、规律和模式学习的模型与方法，这是机器学习的研究范畴。

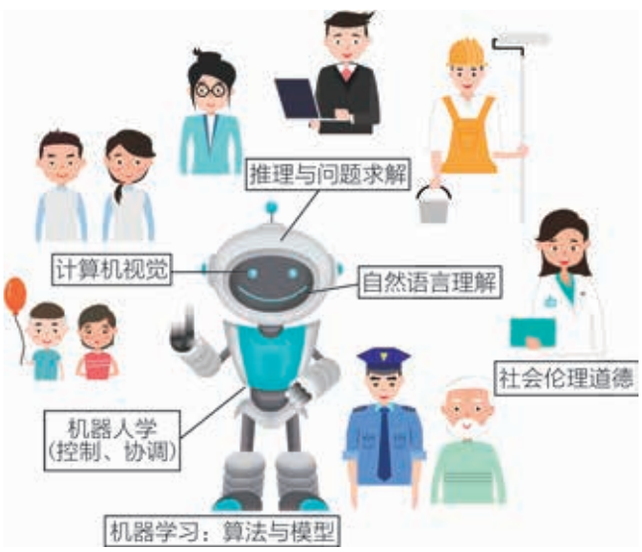


图1.2.4　人工智能所包含内容示意图

如图1.2.4所示，人工智能的核心研究内容，涉及机器学习、推理与问题求解、机器人学、计算机视觉、自然语言理解和社会伦理道德。

1.2.4　人工智能的发展

1955年提出“人工智能”这一概念时，人工智能的“初心”是“复制”人脑功能。但是经过长期发展，人们已经慢慢意识到，在脑科学和神经科学等尚未厘清人脑机理的前提下，科学家难以用机器直接模拟人脑。 物理学家理查德·费曼曾经说过：“不可造者，未能知也。”（What I cannot create, I do not understand）如果把这个说法用于机器“复制”人脑功能这一语境，可以理解为人脑功能的“复制”前提是需要全面了解人脑机理。

除通过计算技术来模拟人类智能以外，当前人工智能发展出现了如下趋势：

（1）脑启发计算。大脑具有感知、识别、学习、联想、记忆和推理等功能，这些功能与大脑结构存在着至今仍然无法确知的对应关系。虽然目前我们尚未完全理解大脑的工作原理，但是来自神经科学的发现可从若干角度影响人工智能的研究。脑启发计算就是通过仿真、模拟和借鉴大脑生理结构和信息处理过程的装置、模型和方法，在结构层次模拟人脑、在器件层次逼近大脑等。

（2）混合增强智能。混合增强智能就是将多种智能混合叠加在一起，从而超越某一智能。如科学家发现壁虎等爬行动物的脚能够稳固地贴在墙上而攀岩走壁，其原理在于“范德华力”（van der Walls）（即分子之间微弱作用力相互组合而形成了惊人力量）。科学家们

围绕壁虎的吸附和脱附功能研制具有先进吸附技术的装置（如航天员固定锚或消防员爬楼手套）或机器人（如电影中飞檐走壁的蜘蛛侠）在极端环境条件下工作，这是生物智能和机器智能的结合。

2014年巴西世界杯期间，28岁的截瘫青年朱利亚诺·平托身穿由肢体辅助装置和神经传感系统组成的“外骨骼机械战甲”为世界杯开出第一球。“外骨骼机械战甲”上装载了神经信号传感器，将朱利亚诺·平托大脑发出的踢球意念信号传输到计算机装置中，计算机装置理解了“踢球”这一意念信号，将其转化为相应指令，驱动液压活塞完成开球动作。这是机器智能辅助人类智能的例子。

在人类智能和其他智能组成的混合增强智能中，人类智能是智能总开关，决定了智能的水平。正如哲学家维特根斯坦所言：“任何人都不能代替你思考，要紧的是深刻地做自己。”

（3）场景人工智能。人工智能算法和模型需要与具体任务和具体场景结合，以便更好地发挥功能，如电商平台中商品推荐、城市道路中“交通指挥大脑”等不同场景。

虽然完全揭示人类智能活动的道路依旧漫长，但是人类会一直在探索智能的道路上不忘初心、矢志前行。“我们必须知道，我们必将知道”，希尔伯特1930年所喊出的与“不可知论主义”相对的话语将激励着人类去探索未知的世界。

拓展链接

新一代人工智能国家规划

2017年7月20日，国务院向全社会公布了《新一代人工智能发展规划》。新一代人工规划将在大数据智能、群体智能、跨媒体智能、混合增强智能和自主无人系统等理论方面取得重大突破，推动人工智能从人工知识表达技术到大数据驱动知识学习，从处理单一类型的数据到跨媒体认知、学习和推理，从追求“机器智能”到迈向人机混合的增强智能，从聚焦研究“个体智能”到基于互联网络的群体智能，从研究机器人到研究智能自主系统。

1.2.5 人工智能开源平台

人工智能已经渗入各行各业，为了快速搭建面向场景的人工智能应用，需要各类人工智能算法的支持。类似于Android手机开源操作系统推动了移动应用发展，当前一些人工智能开源平台也正推动人工智能如火如荼地发展。

科技部2017年11月宣布全面启动国家新一代人工智能重大科技项目时，指定了百度Apollo（面向无人汽车）、阿里阿里云ET城市大脑、腾讯觅影（面向医疗）、科大讯飞语音交互平台四大开放平台，以促进人工智能支撑算法的开放开源，致力于人工智能生态建设，使得人工智能算法能够犹如“水、电和空气”一样便捷地为个人和公司所用。

国际上一些高校和公司也陆续开放了一些人工智能开源工具平台，如加州大学伯克利分校的Caffe、谷歌公司的TensorFlow、英特尔的nGraph、IBM的SystemML、Facebook公司的人工智能硬件平台Big Sur以及人工智能爱好者开发的Theano（基于其开发了Keras、Lasagne和Blocks等深度学习库）、MXNet和Torch等。

本书后续章节要求同学们能够熟悉若干开源平台，在开源平台上搭建神经网络等机器学习模型，进行手写体数字图像识别和图像分类等应用。

实践与体验

Keras平台配置

为了方便后续章节一些实验的进行，需要先学习一种基本实验平台——Keras。Keras是一个高层神经网络API（API是英文Application Programming Interface的简称，即应用程序编程接口），由纯Python编写并可使用TensorFlow或Theano作为后端。Keras为快速实验而生，它能够把设计方案迅速转变为结果，并且具有高度的模块化的特点，支持神经网络计算，可使用CPU或GPU运算。

实践内容：

成功安装和配置Keras。

实践步骤：

1. 在安装Keras前，确保计算机上已经安装好Python。可以通过在cmd控制台输入以下命令验证：python --version，若cmd控制台输出Python的版本号，则认为Python安装成功。

2. 安装Keras。最简易的方法是通过pip安装，在命令行中输入：

```
pip install keras
```

结果呈现：

Keras是否安装成功可以使用以下方法验证：在命令行中输入“python”，然后输入代码“import keras”。如果没有出现错误即可认为Keras安装成功，图1.2.5所示是可能的执行结果。

```
import keras
Using TensorFlow backend.
```

图1.2.5　代码“import keras”的结果呈现

思考与练习

1. 受壁虎的吸附功能的启发，科学家创造了一些具有吸附功能的智能装置。请你列举一些从生物智能获得启发而研制成功的智能装置。

2. 通过各种穿戴设备、人—车协同共驾、脑控或肌控外骨骼机器人、人机协同手术等方式实现了生物智能（主要是人类智能）与机器智能系统的紧密耦合。问：在这种人机混合增强智能形态中，如何理解人类智能是总开关，决定着混合增强智能的高度和深度？

巩固与提高

1. 请给出一个演绎三段论的例子。该三段论中大前提和小前提分别陈述的内容都正确，但是由于大前提和小前提所指向概念的范围不同，导致所得到的推理结果是错误的。

2. 根据图 1.1.1 示例过程，请给出图灵机对“5+10−6−3”这一问题的计算过程，并判断该问题是否为图灵可停机问题（即该问题是否为可计算问题）。

3. 请从基础理论研究突破（如何结合神经科学与认知科学研究进展）、算法模型改进（如何建立小数据、大任务等模型）和任务明确的场景人工智能实现（面向养老、健康、教育和环保等领域创新应用）等角度来讨论如何保证人工智能健康发展，让人工智能拥抱实体经济、服务社会、造福人类。

项目挑战

构建人工智能发展脉络全景图

人工智能被提出的初期是指以机器为载体来实现人类智能或生物智能。这里涉及三个基本问题：（1）模拟人类智能的机器载体是什么？从本书介绍可知图灵机可机械式完成“所有可计算任务”。（2）如何通过算法来模拟人类智能？人工智能需要通过一些算法来实现人类的某些行为。（3）哪些人类智能在目前情况下是无法实现的？

项目任务

设计一张人工智能发展脉络的全景图，向不了解人工智能的人介绍智能实现的载体、人类智能的组成部分以及人工智能的能与不能等问题。

具体要求如下：

1. 图文并茂地展示对计算载体、智能算法和智能极限的思考。
2. 将图灵机与现代计算机对比，分析其成为现代计算机理论模型的原因。
3. 列举若干位与人工智能研究有关的图灵奖获得者的成就。
4. 从图 1.2.4 中的角度来列举人工智能研究中代表性的算法。
5. 分析现有人工智能的局限性及相应对策。

过程与建议

1. 人工智能发展脉络全景图的构成元素

为了构建人工智能发展脉络全景图，建议从以下方面展开思考：

（1）现代计算机理论模型图灵机的诞生缘由。围绕可计算这一问题，认真考虑为什么图灵机成为现代计算机的理论模型，对比图灵机模型与现代计算机之间的异同。

（2）人工智能研究范畴所涉及的算法功能。人工智能研究涉及计算机视觉、自然语言理解、推理与问题求解、机器学习等，从这些方面列举人工智能能实现的若干功能。

（3）人工智能局限性与智能提升途径。计算机程序能够实现人类哪些智能，不能实现人类哪些智能？是否存在克服这些困难的可行手段（如人—机相互合作）？

2. 绘制人工智能发展脉络全景图

从计算载体、智能算法、智能极限的角度，绘制人工智能发展脉络全景图，该全景图要清晰地描述如下内容：实现人工智能载体的计算机理、人工智能研究范畴中代表功能、

人工智能发展局限及其对策、人工智能领域获得图灵奖的学者。

3. 撰写项目报告

基于以上工作，根据所绘制的人工智能发展脉络全景图，撰写项目报告，应包括如下内容：

（1）研究背景与目标。
（2）人工智能发展脉络全景图。
（3）可计算思想起源。
（4）图灵机模型。
（5）人工智能代表功能。
（6）人工智能领域图灵奖获得者及其主要成就。
（7）人工智能的局限性及可能提升手段。

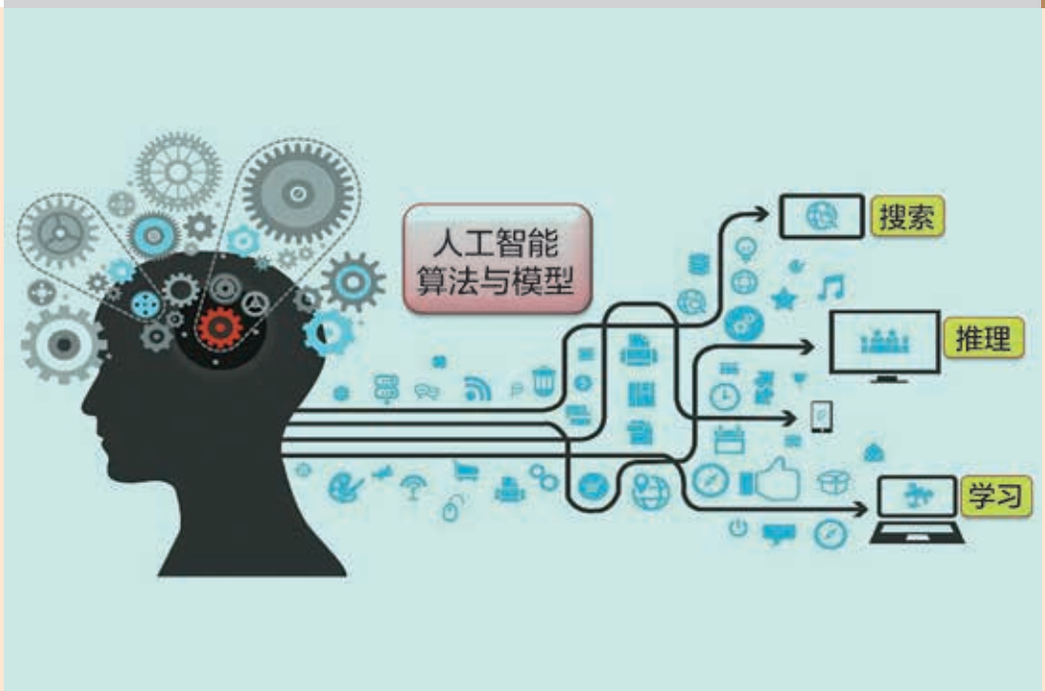
评价标准

请根据项目实施的过程、效果以及成果展示交流的结果，对自己完成项目的情况进行客观的评价，并思考后续完善的方向。把评价结果和完善方案填写在下面的表格中。

评价条目	说明	评分（1～10分）	评分主要依据阐述	后续完善方向
项目理解	在项目开展和分析报告中，展现出对项目目标、内容与任务的正确理解			
人工智能发展脉络全景图	通过时间轴来组织内容，内容紧凑、布局合理			
计算载体	分析图灵机和现代计算机的异同			
人工智能代表性人物	列出人工智能领域获得图灵奖的代表学者及其主要成就			
人工智能的功能和局限性	列出若干功能描述，合理分析局限性			
展示交流	分析正确，展示经过精心准备，表达清晰、便于受众理解			

第二章 2

智能之源：算法与模型



从实现角度而言，人工智能是运用数学模型来设计求解算法，然后运行程序代码得到结果。推理、搜索和学习是人工智能实现所采用的主要方向。

逻辑推理是从形式化角度对人脑认知的一种模拟，如命题逻辑和谓词逻辑等。在逻辑推理中，所有的判断（命题）均被表达为规则，于是可从规则出发，推导出新的判断（命题），专家系统就是以逻辑推理为核心来模拟人类某一方面智能的系统。

另外，也可从非形式化角度来实现智能模拟：（1）以搜索为核心的问题求解方法。在这种方法中，给定一个待求解目标，通过所设定的搜索模型来求解问题答案，如启发式搜索算法。（2）以规则为核心的决策树方法。该方法将若干规则组合成树状结构，然后一步步应用不同规则来解决求解问题，分而治之，找到问题答案。（3）以数据为核心的机器学习方法。该方法须从已有数据学习蕴含模式，如在输入数据和输出结果之间建立关联关系的回归拟合模型、从已知的输入数据去计算结果概率大小的贝叶斯模型、对输入数据层层学习的神经网络模型等。

随着脑观测和脑机接口等技术的进步，可将人的作用或认知模型引入人工智能系统中，形成混合增强智能形态，这也是人工智能未来发展的重要模型。

问题与挑战

- 为了仿真大脑智能活动，我们首先需要观测大脑中海量神经元和神经突触的复杂活动。人类目前在大脑观测方面取得了较大进展。一旦能够完全观测清楚大脑运行机制，是否意味着仿真大脑的目标已可实现？

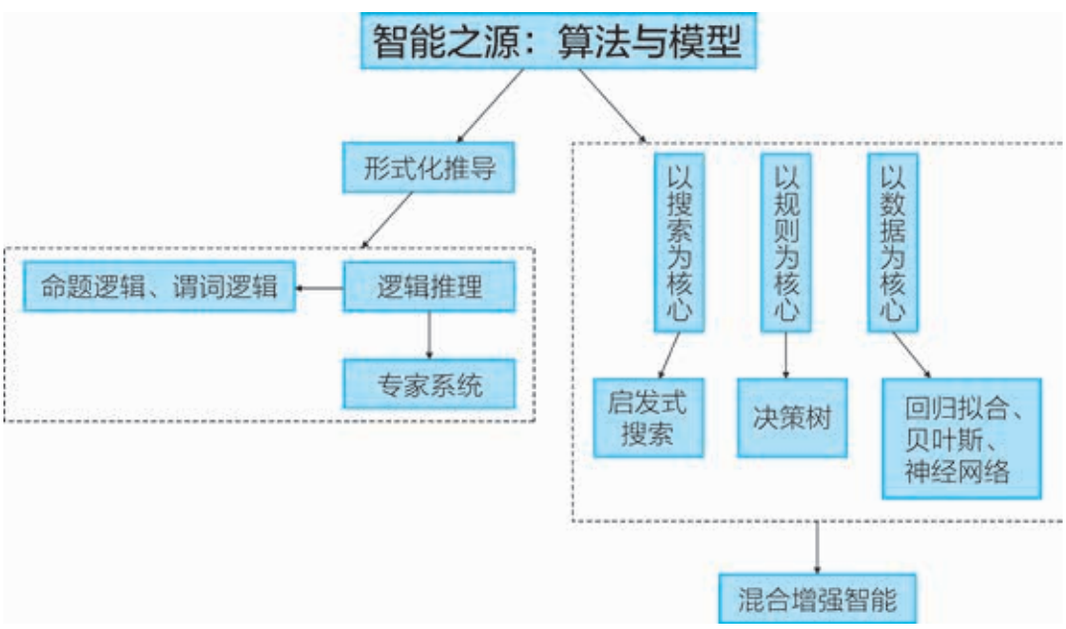
- 现有智能算法或系统一般是“依葫芦画瓢”，难以“举一反三”，即算法通过模仿，而非完全理解去完成某一任务。“授之以鱼，不如授之以渔”，为训练完成某任务的一个智能系统，不仅要给它大量的数据和规则以便其模仿，还需要抽象出该任务的内在机理以便智能算法能完成该任务。思考“授之以鱼”和“授之以渔”两种人工智能实现理念的不同。

- 互联网上存储了大量事实型知识（如每年农历八月十五，钱塘江涌潮最大），在这种情况下，人类是否已无必要记住事实型知识？从智能模拟角度进行思考。

学习目标

1. 了解脑认知机理和逻辑推理基本概念。
2. 掌握搜索求解、决策树分类和回归分析等基本算法。
3. 了解神经网络基本模型、学习过程和搭建方法。

内容总览



2.1 类脑计算

类脑计算（Brain-like Computation）是指仿真、模拟和借鉴大脑神经系统结构和信息处理过程的装置、模型和方法，其目标是制造类脑计算机或实现类脑智能。

与经典人工智能符号主义、连接主义、行为主义以及机器学习的统计主义这些技术路线不同，类脑计算采取如下仿真主义：结构层次模仿脑、器件层次逼近脑（神经形态器件替代晶体管）、智能层次超越脑（智能主要通过自主学习获得而不是人工编程定义）。

经过百万年演化而成的人脑或许是世界上最复杂、最精密的机器，它承载了人类所有智能活动（注意、学习、记忆、直觉、顿悟和决策等）。人脑的核心结构由百亿个神经元及百万亿神经突触构成。如图2.1.1所示，每个神经元通过神经突触“接收”从其他神经元传递过来的信息，再将加以处理后的信息通过神经突触传递给其他神经元。

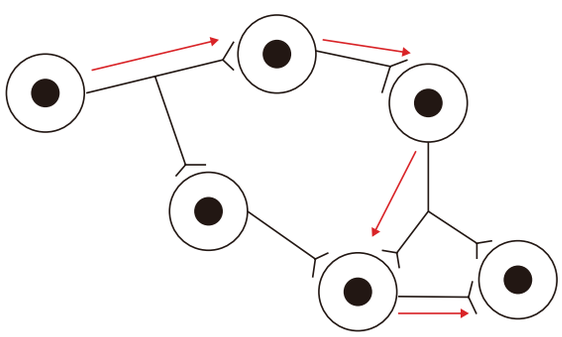


图2.1.1　神经元间的信息传递

这样，外界感官信息（视觉、听觉、嗅觉、味觉、触觉）以复杂方式经过大脑分析、处理，形成感知和认知。

“工欲善其事，必先利其器。”为了仿真大脑对外界信息的处理机制，首先要观测信息在大脑神经元和神经突触之间的“传递”。目前，光、电、磁、声等观测与调控技术为大脑观测提供了有效手段，打开了理解大脑、模拟大脑、连接大脑的大门。当前较为成熟的大脑观测方式有CT（计算机断层扫描）、MRI（磁共振成像）、PET（正电子发射计算机断层扫描）和fMRI（功能性磁共振成像）等。

记忆是大脑智能的基础。神经科学研究发现，人脑有瞬时记忆、工作记忆（有些文献也称为短时记忆）和长期记忆三种形式的记忆体（如图2.1.2）。

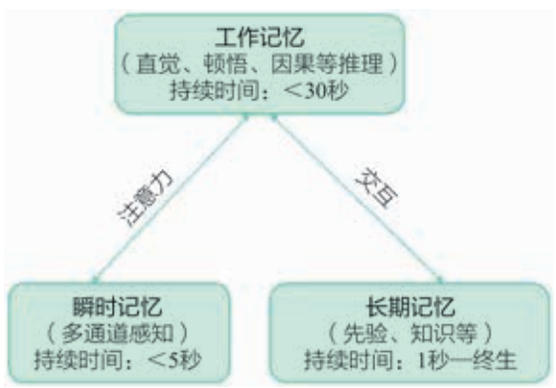


图2.1.2　三种记忆体及联系

瞬时记忆用来感知外界信息，如通过“眼观六路、耳听八方”从外界环境中不断感知信息。只有那些被注意了的数据才会送入工作记忆体，直觉、顿悟和推理等智能活动就基于这些数据在工作记忆体中开展。只是，在进行这些智能活动时，人脑会从长期记忆体中唤醒和激活与这些被处理数据相关

的信息（如历史记忆、类似例子等），和当前数据综合在一起进行处理。

古语云，“弦外之音”“画外之意”。这从一个侧面描述了人脑在认知理解中不仅依赖当前数据，而且从长期记忆中唤醒了类似信息，将这些信息与当前数据综合在一起，进行加工和处理。

许多科学家认为，工作记忆在大脑中的海马体形成与储存，长期记忆在大脑皮层长期储存。

可见，如果设计算法对上述三种记忆体之间的交互和融合进行模拟，则可建立起“类脑计算”模式，更好地模拟大脑功能，从而完成智能处理任务。

从上述描述可知，人脑学习与记忆功能由如下三部分组成：将事件编码进入神经网络；将编码好的信息储存为长期记忆；在需要的时候从长期记忆中调出相关信息。这个过程如图2.1.3所示。

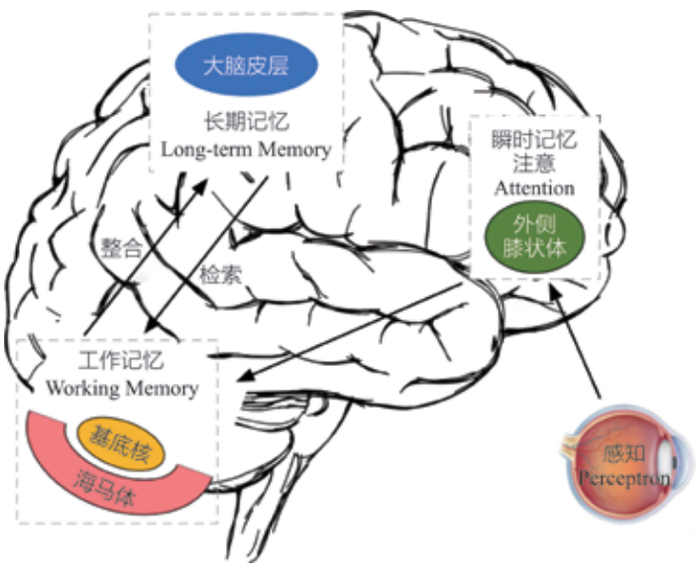


图2.1.3　大脑皮层中长期记忆和海马体中工作记忆的示意图

因此，理解人类认知中记忆等神经基础是促进人工智能取得长足发展的一个具有吸引力的目标。从对大脑功能的观测理解中，来获得人工智能发展所需的启发机理，是目前脑科学与人工智能交叉领域的活跃方向。但是，科学家至今对人类认知功能如何从复杂动态（时空演变）的大脑神经结构中产生依然没有形成较为完整的认识，这是一个充满挑战的征途。

拓展链接

飞秒成像（Femto Photograph）

1飞秒是1秒的一千万亿分之一（即10^{-15}秒），即使是每秒飞行30万千米的真空中的光，在1飞秒内，也只能走300纳米（相当于人体头发直径的二百分之一）。为了更好地观测大脑神经元和神经突触活动，科学家正在研制“飞秒成像”光学设备，希望1秒钟能够拍摄万亿张左右的图像，以实现对大脑细微活动的观测，从而反演大脑功能和机理，促进人工智能进一步发展。

《庄子·天下》中写道：“至大无外，谓之大一；至小无内，谓之小一。”2500年前希腊哲学家对物质的组成问题争论不休。2018年，科学家通过电子显微镜像素阵列探测器获得了0.000000000039米的电子显微镜成像分辨率。

看得快和看得清为探究人脑机理奠定了实验观测基础。

2.2 逻辑推理

逻辑学是探索、阐述和确立有效推理原则的学科。一般而言，逻辑学是用数学方法来研究关于推理和证明等问题的学科。早在17世纪，德国数学家莱布尼茨就曾经设想创造一种“通用的科学语言”，以便能像数学应用公式一样来对推理过程进行计算，进而得出正确的结论。

人工智能研究的一个核心问题是使用计算机去模拟人类思维活动。人类思维活动一个重要功能是逻辑推理，即通过演绎和归纳等手段对已有观测现象进行分析，加以判断。在人工智能发展初期，脱胎于逻辑推理的符号主义人工智能（symbolic AI）是人工智能研究的一种主流学派。

在符号主义人工智能中，所有概念均可以通过人类可理解的“符号”及符号之间的关系来表示。若使用符号A来表示对象概念、*IsCar*()来表示某个对象是否为“汽车”，则*IsCar*(A)表示“A是一辆轿车”。*IsCar*(A)由对象A和*IsCar*()两部分所构成。若A是轿车，则*IsCar*(A)为正确描述，否则是错误描述。

符号主义人工智能方法基于如下假设：可通过逻辑方法来对符号及其关系进行计算，实现逻辑推理，辨析符号所描述内容是否正确。本节将重点介绍命题逻辑（propositional logic）和谓词逻辑（predicate logic），以及与它们相关的逻辑推理规则。同时，还将介绍逻辑推理在人工智能领域应用的重要产物——专家系统（expert system）、知识图谱和常识推理等内容。

拓展链接

墨子的逻辑学思想

墨翟，尊称为墨子，被认为是东方逻辑学的奠基人。墨子提出了名、辞、说三种基本思维形式和由故、理、类三物构成的逻辑推理；也提出了一些几何思想，如“平，同高也”（两平行线或两平行平面间距离处处相等），“圆，一中同长也”等。

2.2.1 命题逻辑

命题逻辑是一套应用形式化规则对以符号表示的描述性陈述进行推理的系统。在命题逻辑中，一个或真或假的描述性陈述被称为原子命题，若干原子命题可通过逻辑运算符来构成复合命题。通常使用小写字母（如p，q，r，s，t等）来表达命题。

下面给出了五个陈述，请判断它们的真假。

p：北京是中国的首都

q：13能被6整除

r：$x<8$

s：存在最大的素数

t：$m^2 \geqslant 0$（m是实数）

可以看到，陈述p和陈述t是真命题，陈述q和陈述s是假命题。陈述r的真假取决于x的取值，因此r不是一个命题。

任何一个命题或为真或为假（二者必取其一）。

如果已知若干命题，那么可通过命题联结词（connectives）对已有命题进行组合，得到新命题。这些通过命题联结词得到的命题被称为复合命题（compound proposition）。

下面介绍五种主要的命题联结词：

与（and）：表示命题合取（conjunction），即“p且q”，用符号∧来表达。

或（or）：表示命题析取（disjunction），即“p或q”，用符号∨来表达。

非（not）：表示命题否定（negation），即“非p”，用符号¬来表达。

条件（conditional）：表示命题蕴含（implication），即“如果p那么q”，其中p被称为假设或前提，q被称为结论，用符号→来表达。

双向条件（biconditional）：表示命题双向蕴含（bi-implication），即“p当且仅当q”，用符号↔来表达。

例如，假设p表示命题“今天下雨”、q表示命题“户外表演会被取消”，通过条件联结词可从命题p和q得到一个新命题：$p \rightarrow q$，表示“如果今天下雨，那么户外表演会被取消”。

无论是原子命题，还是复合命题，其或为真或为假。命题的真假结果被称为真值（truth value）。

对于单个命题，较易判断其为真（用T表示）或为假（用F表示）。而对于复合命题，由于涉及命题联结词，判断其真假较为困难，可借助真值表（truth table）来判断其真假。具体而言，真值表通过穷举复合命题中原子命题的真假组合，通过命题联结词进行逻辑运算来判断复合命题真假。

下面以“与”和“条件”这两个命题联结词为例，来展示如何通过真值表判断复合命题真假。

“与”命题联结词连接两个命题p和q，当其中有一个命题为假时，则复合命题$p \wedge q$为假。由此可得“与”命题联结词所构成复合命题的真值表，如表2.2.1所示。

表2.2.1 “与”命题联结词真值表

p	q	$p \wedge q$
T	T	T
T	F	F
F	T	F
F	F	F

表2.2.2　给定具体原子命题前提下“与”命题联结词真值表

p	q	$p \wedge q$
太阳从东边升起	偶数能被2整除	真命题
太阳从东边升起	偶数不能被2整除	假命题
太阳从西边升起	偶数能被2整除	假命题
太阳从西边升起	偶数不能被2整除	假命题

表2.2.2给出了原子命题为真或为假时，复合命题的真假情况。可以看出，两个原子命题通过“与”命题联结词进行组合，只有两个原子命题都为真命题时，所得到的复合命题才为真。

用“条件”命题联结词连接两个命题p和q，当p为真且q为真或p为假，所构成的复合命题$p \rightarrow q$为真。这样，可得到如表2.2.3所示的“条件”命题联结词所构成复合命题的真值表。

表2.2.3　“条件”命题联结词真值表

p	q	$p \rightarrow q$
T	T	T
T	F	F
F	T	T
F	F	T

表2.2.4　给定具体原子命题前提下“条件”命题联结词真值表

p	q	$p \rightarrow q$
太阳从东边升起	偶数能被2整除	真命题
太阳从东边升起	偶数不能被2整除	假命题
太阳从西边升起	偶数能被2整除	真命题
太阳从西边升起	偶数不能被2整除	真命题

从表2.2.4可知，只有命题p为真命题、命题q为假命题时，$p \rightarrow q$为假命题。其他任何情况下，$p \rightarrow q$均为真命题。

对于其他的命题联结词，可以通过上述类似过程得到其所构成复合命题所对应真值表，这些命题联结词的真值表是进行逻辑运算的基础。

在逻辑推理中有一个概念叫逻辑等价（logical equivalent）。给定命题p和q（p和q一般是复合命题），如果p和q在所有情况下都具有同样真假结果，那么p和q在逻辑上等价，用≡来表示，即$p \equiv q$。

逻辑等价为命题进行形式转换带来了可能，基于这些转换不再需要逐一列出p和q的真值表来判断两者是否在逻辑上等价，而是可直接根据已有逻辑等价公式来判断p和q在逻辑上是否等价。

下面列出了一些重要的逻辑等价关系：

同一律（identity law）：$p \wedge T \equiv p$，$p \vee F \equiv p$。

支配律（domination law）：$p \vee T \equiv T$，$p \wedge F \equiv F$。

幂等律（idempotent law）：$p \wedge p \equiv p$，$p \vee p \equiv p$。

双重否定（double negation law）：$\neg(\neg p) \equiv p$。

德摩根律（de morgan's law）：$\neg(p \wedge q) \equiv \neg p \vee \neg q$，$\neg(p \vee q) \equiv \neg p \wedge \neg q$。

拓展链接

“条件”命题联结词中前提为假时命题真假取值

条件命题联结词所构成的复合命题描述了“如果p那么q”的含义，记为$p \rightarrow q$，其中p为假设或前提，q为结论。如果命题p为假，那么$p \rightarrow q$为真，无论命题q是否为真或假。比如，复合命题“太阳从西边升起→偶数不能被2整除”，因为前提“太阳从西边升起”是错误的，即使结论是错误的，该复合命题仍然是正确的。

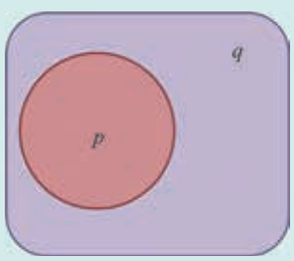


图2.2.1　条件命题联结词中的“前提”和“结论”包含关系

“如果p那么q”（$p \rightarrow q$）定义的是一种蕴含关系（即充分条件），也就是命题q包含着命题p（p是q的子集，如图2.2.1所示），p不成立相当于p是一个空集，空集可被其他所有集合所包含，因此当p不成立时，“如果p那么q”永远为真。

思考与练习

给定命题p，q和z，通过“与”“或”“非”命题联结词得到复合命题$(p \vee q) \wedge \neg z$，该复合命题可以是真命题吗？如果该复合命题为真命题，p，q和z的真假取值是什么？

2.2.2　谓词逻辑

在谓词逻辑中，原子命题可分解成个体和谓词。个体是可独立存在的事或物，谓词则是用来刻画个体具有某些性质。

考虑下面两段陈述：

I：如果下雨，当天户外演出会被取消。昨天下雨了。昨天户外演出被取消。

II：$x < 11$

这两段陈述如果用命题逻辑来表达，存在如下局限：I中三句话可视为三个原子命题，

但三个原子命题之间所存在的关联性无法得到体现；II 的真假无法判断。

为了刻画命题内在关联性，以及对内部包含变量的陈述进行逻辑表达，就需要比命题逻辑更为强大的工具——谓词逻辑，又称为一阶逻辑（first-order logic）。

下面将介绍谓词逻辑中两个核心概念：谓词（predicate）和量词（quantifier）。

一个陈述句可分为两个部分，分别是个体和谓词。谓词在形式上就是陈述句中刻画关系的部分，如可将“$x<11$”这句陈述分解为个体（x）和谓词（<11）两个，其中“<11”刻画了“小于 11”这种关系。

一般用大写字母（如 P，Q，R 等）表示谓词，如“$x<11$”记为 $P(x)$，其刻画了一个数是否小于 11 的关系。将个体从命题中独立出来，既能建立命题之间的关联，又能够容忍不确定性的存在。将关系从命题中独立出来，则可以在一个命题中刻画若干个体之间的关系或刻画多种关系。

再如，用 $P(x_1, x_2)$ 表示“$x_1>x_2$”这一命题，则 $P(0, 3)$ 表达的是“$0>3$”。显然，$P(0, 3)$ 是一个假命题。

谓词逻辑中个体有一定的取值范围，这个范围就是个体域（domain）。除了给个体指定一个赋值外，还有两种非特指表达方式：“某些”和“所有”。量词就是在谓词逻辑中对个体进行非特指表达的工具，常用的两种量词是存在量词（existential quantifier）和全称量词（universal quantifier）。

存在量词表示“存在某个或某些个体”满足谓词逻辑所述关系，用符号∃表达。全称量词表示“每一个个体”满足谓词逻辑所述关系，用符号∀表达。

假设个体域为所有的自然数，用 $Q(x)$ 来表示“x 的平方大于等于 0”（$x^2\geqslant 0$）这样的命题。通过引入存在量词，可以定义 $\exists xQ(x)$ 这一命题，其表达了“存在一个 x，其平方大于等于 0”；通过引入全称量词，可以得到 $\forall xQ(x)$ 这一命题，其表达了“所有自然数的平方都大于等于 0”。显然，$\exists xQ(x)$ 和 $\forall xQ(x)$ 都是真命题。

表 2.2.5　谓词逻辑中的存在量词和全称量词（个体域为整数域）

谓词逻辑 $Q(x)$	$\exists xQ(x)$ 为真命题时取值	$\forall xQ(x)$
$Q(x)$：$x^2\leqslant 0$	$x=0$	假命题
$Q(x)$：$x+10\leqslant 10$	$x=-6$	假命题

与命题逻辑一样，可在谓词逻辑中应用命题联结词，组合得到新的命题。

思考与练习

1. 假设谓词逻辑 $P(x)$ 表示“$x^2-4\geqslant 0$”这一命题，请在个体域为整数域的情况下，判断 $\exists xP(x)$ 和 $\forall xP(x)$ 为真或为假时的取值。

2. 将陈述“在所有三角形中，任意两条边之和一定大于第三条边”用谓词逻辑表述。

2.2.3 逻辑推理规则

对客观世界按照命题逻辑或谓词逻辑进行形式化描述，再基于一定规则，从已有命题出发，得到一些结论，这个过程称为逻辑推理。

逻辑推理的一般过程如下：

$$\frac{\begin{array}{c}\text{前提 } 1\\ \vdots\\ \text{前提 } n\end{array}}{\therefore \text{结论}}$$

从n个前提出发，推导出结论，即$(p_1 \wedge \cdots \wedge p_n) \rightarrow c$，其中$p_i(i=1,\cdots,n)$是前提，$c$是结论。

如下给出一些重要的推理规则：

（1）肯定前件（modus ponens）：$\dfrac{\begin{array}{r}p\\ p \rightarrow q\end{array}}{\therefore q}$

（2）否定后件（modus tollens）：$\dfrac{\begin{array}{r}\neg q\\ p \rightarrow q\end{array}}{\therefore \neg p}$

（3）假言三段论（hypothetical syllogism）：$\dfrac{\begin{array}{r}p \rightarrow q\\ q \rightarrow r\end{array}}{\therefore p \rightarrow r}$

（4）析取三段论（disjunctive syllogism）：$\dfrac{\begin{array}{r}\neg p\\ p \vee q\end{array}}{\therefore q}$

（5）消解（resolution）：$\dfrac{\begin{array}{r}p \vee q\\ \neg p \vee r\end{array}}{\therefore q \vee r}$

（6）全称量词消去（universal instantiation）：$\dfrac{\forall x P(x)}{\therefore P(c)}$

（7）存在量词消去（existential instantiation）：$\dfrac{\exists x P(x)}{\therefore P(c) \text{ for some element } c}$

下面通过两个具体例子来理解逻辑推理的过程。

●●● 例 1

命题p：小高帮助小王写代码

命题q：小王今天完成代码编写工作

命题r：小王晚上可以早睡

已知前提：

$p \rightarrow q$//如果小高帮助小王写代码，那么小王今天完成代码编写工作

$q \rightarrow r$//如果小王今天完成代码编写工作，那么小王晚上可以早睡

$\neg r$//小王晚上没有早睡

证明：$\neg p$//小高没有帮助小王写代码

下面给出证明过程：

（1）根据假言三段论规则，可以得到$p \rightarrow r$（如果小高帮助小王写代码，那么小王晚上可以早睡）；

（2）根据否定后件规则，可以得到$\neg p$（小高没有帮助小王写代码）。

例2

定义命题：

$P(x)$：x为正整数；$Q(x, y)$：x大于y；$R(x, y)$：x的平方大于y的平方

已知前提：

$P(a)$//a为正整数

$P(b)$//b为正整数

$\neg R(a, b)$//a的平方不大于b的平方

$\forall x \forall y(P(x) \wedge P(y) \wedge Q(x, y) \rightarrow R(x, y))$//对于任意的正整数$x$和$y$，如果$x$大于$y$，那么$x$的平方大于$y$的平方

证明：$\neg Q(a, b)$//a小于b

下面给出证明过程：

（1）根据全称量词消去规则，得到$(P(a) \wedge P(b) \wedge Q(a, b)) \rightarrow R(a, b)$（对于正整数$a$和$b$，如果$a$大于$b$，那么$a$的平方大于$b$的平方）；

（2）根据否定后件规则，得到$\neg(P(a) \wedge P(b) \wedge Q(a, b))$（$a$和$b$是正整数且$a$大于$b$是不成立的）；

（3）根据德摩根律，得到$\neg P(a) \vee \neg P(b) \vee \neg Q(a, b)$（或者$a$不是正整数，或者$b$不是正整数，或者$a$小于$b$）；

（4）根据析取三段论规则，得到$\neg Q(a, b)$。

从如上两个例子可看出，通过灵活运用命题的逻辑等价关系和逻辑推理规则，就可从前提出发证明一些结论或说明某些结论无法从现有前提得出。

由于运用推理规则和等价关系顺序不同可产生不同的推理过程，因此推理的中间过程并非唯一。

拓展链接

吴文俊与机器定理证明

吴文俊是中国著名数学家，他通过研究中国古代数学并汲取其特色，创立了机器证明几何定理的“吴方法”，开创了数学机械化道路，并因此在2000年与袁隆平一起获得首届国家最高科学技术奖。所谓数学机械化，就是在运算或证明过程中，每执行一步操作后，都有一个确定的、必须选择的下一步操作，这样机器沿着一条有规可循道路进行机械操作，直到得到结论。吴文俊曾明确指出：中国古代数学是一种机械化数学，强调构造性、算法化。

思考与练习

利用消解规则证明如下命题（这两个命题分别称为构造性二难和破坏性二难）：

$$\begin{array}{r} p \vee r \\ p \rightarrow q \\ \underline{r \rightarrow s} \\ \therefore q \vee s \end{array} \qquad \begin{array}{r} p \rightarrow q \\ r \rightarrow s \\ \underline{\neg q \vee \neg s} \\ \therefore \neg p \vee \neg r \end{array}$$

（提示：要用到“条件”联结词的逻辑等价关系 $p \rightarrow q \equiv \neg p \vee q$）

2.2.4 专家系统

逻辑推理在人工智能领域最重要的应用之一，就是专家系统。所谓专家系统，简单来讲就是能完成某一特定专家任务的计算机系统。

专家是具备特定领域丰富知识和精通特定技能的一类人才，因此能够模拟人类专家的计算机系统同样应具备这些特点。具体来说，专家系统需要具备面向特定领域的专家知识，并能够基于这些领域知识，以类似专家的思维进行推理并解决实际问题。

拓展链接

早期专家系统DENDRAL

最早的专家系统雏形是斯坦福大学的费根鲍姆（E.A. Feigenbaum，1994年图灵奖获得者）和化学家勒德贝格（J. Lederberg）在1965年合作研制的DENDRAL系统。该系统通过基于知识表达与推理规则的人工智能方法，可根据给定有机化合物的分子式和质谱图，从几千种可能的分子结构中挑选出一个正确的分子结构，帮助有机化学家鉴定未知有机分子。DENDRAL理念为后续专家系统发展提供了借鉴。1977年，费根鲍姆在第五届国际人工智能会议上提出了“知识工程”（knowledge engineering）这一概念，进一步推动了专家系统的发展。

20世纪70年代开始，专家系统逐渐被人们接受。许多专家系统被相继研发出来，如医药专家系统MYCIN。MYCIN会询问用户一系列判断题和一些描述性问题，逐步通过用户回答来进行规则搜索和推理，进而依据概率大小给出一个可能导致该疾病的细菌列表，并据此推荐治疗方案，如推荐抗生素及其剂量等。

专家系统的典型结构如图2.2.2所示，其由人—机交互界面、知识库和推理机三个主要部分构成。专家系统工作流程如下：

（1）领域专家通过人—机交互界面，将领域内的专业知识（如医学）转化为机器可读（machine readable）形式，存入知识库中。当然，专家系统也通过人—机交互界面输出答案。

（2）用户向专家系统输入数据，推理机根据输入数据，应用推理规则进行推理，得到

结果，将结果通过人—机交互界面反馈给用户。

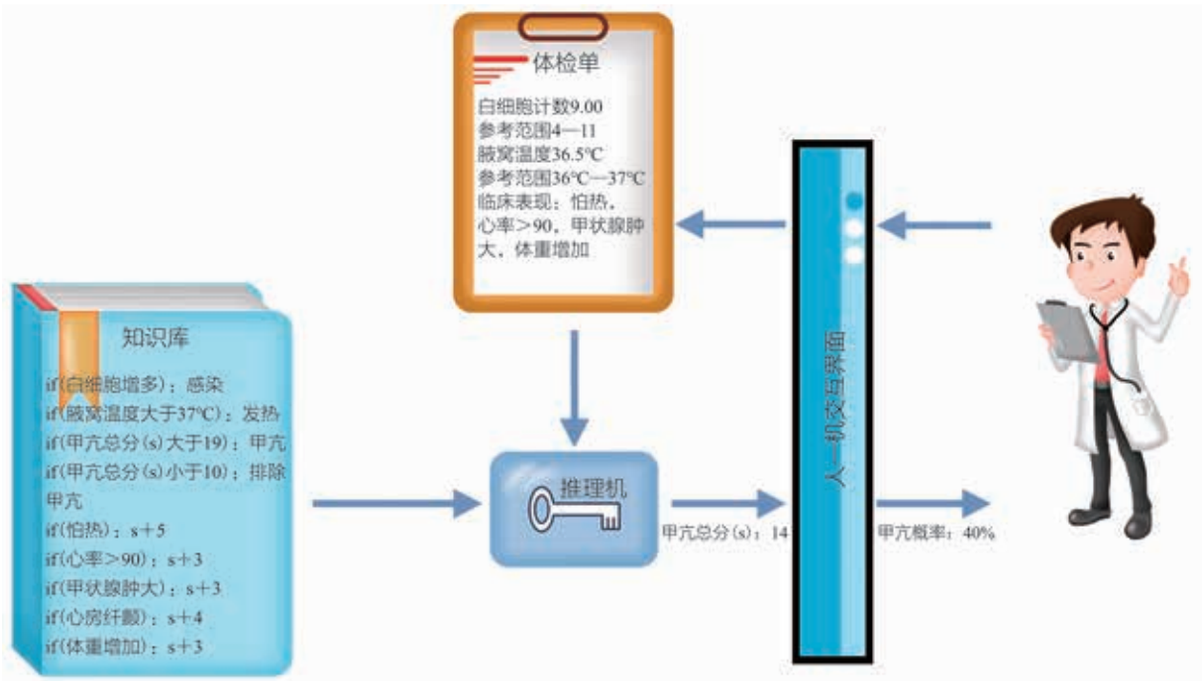


图2.2.2　专家系统组成结构

在图2.2.2中，用户提交体检的相关信息，专家系统基于知识库中存储的规则，推理出用户患甲亢的概率为40%。

知识库存储了问题求解所需的领域知识，如事实、规则和其他信息。知识的表示形式可多种多样，其中主要表示形式就是由“条件”命题联结词所构成的蕴含规则。知识库中所包含知识的质量好坏和数量多少决定了专家系统的水平，如果所输入的问题在知识库中查询不到对应的知识，那么专家系统也无计可施。

推理机是专家系统的核心，根据用户信息，应用推理规则来获取问题的解决方案。

专家系统中的知识库与推理机相互独立，这样只要将一个领域的知识更换为另一领域的知识，在不改变推理机的情况下，就可以在一个领域专家系统的基础上快速开发另一领域专家系统。

2.2.5　知识图谱

1. 知识图谱的意义

从前面内容可知，人工智能在推理过程中需要使用结构化知识。图2.2.3给出了一段描述语句，你能从这段语句中读出什么有意义的信息？

> 北京故宫是中国明、清两代的皇家宫殿，位于北京中轴线的中心，严格地按《周礼·考工记》中“左祖右社，前朝后市”的帝都营建原则建造，是中国古代宫廷建筑之精华。北京故宫以太和殿、中和殿、保和殿三大殿为中心，占地面积72万平方米，建筑面积约15万平方米，有大小宫殿七十多座，房屋九千余间，是世界上现存规模最大、保存最为完整的木质结构古建筑之一。

图2.2.3　北京故宫的描述语句

这段描述北京故宫的语句中包含了故宫的位置、建造原则、三大殿构成、建筑面积等有意义的信息。如果能够将这些信息组织起来，形成结构化知识，就能实现“从字符串信息”到“结构化信息”的转换，这就是知识图谱（Knowledge Graph）所要完成的“从数据到知识”的任务。

可以设想，一旦有了这些结构化知识，如（故宫，三大殿组成，太和殿—中和殿—保和殿），（故宫，是，明、清两代的皇家宫殿），（古代帝都，建造原则，左祖右社和前朝后市），（左祖右社和前朝后市，来自，《周礼·考工记》）等，就可知晓“故宫营建原则是什么？”“来自什么文献？”等问题的答案，而不是通过匹配式搜索手段从描述句子中“寻找”答案。

具体而言，知识图谱是刻画概念之间关联关系的网络，是一种针对海量数据的有效组织方式和利用手段。现有代表性知识图谱包括WordNet、DBpedia、Freebase以及百度百科和维基百科等。

2. 知识图谱的构成

图2.2.4给出了有关水利信息的一个知识图谱片段。

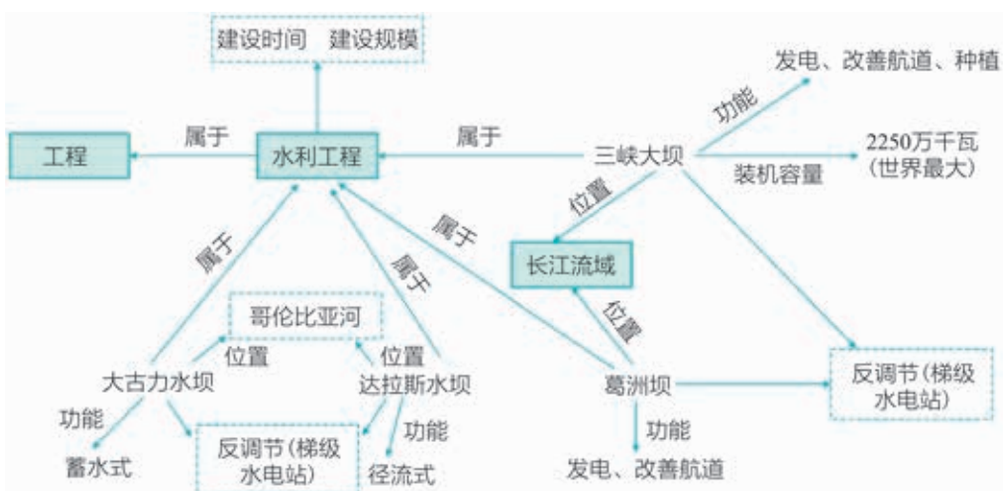


图2.2.4　水利工程领域知识图谱片段

下面介绍知识图谱的构成元素。

（1）概念。知识图谱刻画了概念之间的关联关系，如图2.2.4中工程和水利工程是具有层次化关系的概念、三峡大坝和葛洲坝是水利工程的具体表现。知识图谱将不同概念以图的形式组织在一起。

（2）属性。为了解释某个概念，往往需要刻画其内涵，这个内涵可通过属性来描述。如水利工程具有建设时间和建设规模等属性、三峡大坝具有装机容量及位置等属性。在知识图谱中，如果某个概念和另外一个概念存在上下层次化关系，那么下位概念可继承其上位概念的所有属性（如三峡大坝可继承水利工程所有属性）。

（3）关系。概念之间具有某些关系，如三峡大坝和葛洲坝由于都位于长江流域，它们

之间具有反调节关系（梯级水电站）。关系将两个概念联结到一起。

根据图2.2.4所示的知识图谱内容，如果当推理机“知道”两个水电站位于同一个水域，并且它们具有反调节关系，那么它可以推理出大古力水坝和达拉斯水坝也应该具有反调节关系，即使大古力水坝和达拉斯水坝之间具有的反调节关系在知识图谱中开始并没有定义。

3. 知识图谱的构建

知识图谱需要从数据中提取结构化信息，以图的形式将这些信息组织起来。一种构建知识图谱的方法是将百科词条作为概念，把该词条中信息框（infobox）信息作为概念属性，如YAGO、DBpedia和Freebase等就采取了这样的构建方法。图2.2.5中给出了百度百科中“西湖”这一词条中的信息框信息。

中文名称	杭州西湖	湖泊面积	6.39平方千米
外文名称	West Lake Cultural Landscape of Hangzhou	平均深度	2.28米
地理位置	中国浙江杭州	库容量	1429万立方米
气候类型	亚热带季风气候	湖岸线长度	15千米
占地面积	约6.4平方千米	经纬度	30°14′45″N 120°8′30″E
开放时间	全天开放	名　诗	《饮湖上初晴后雨》等
景点级别	国家5A级风景名胜区	湖中一山	孤山
门票价格	免费	湖中二塔	保俶塔，雷峰塔
著名景点	西湖十景 新西湖十景 三评西湖十景	湖中三岛	小瀛洲，湖心亭，阮公墩
遗产名录	世界文化景观遗产	湖中三堤	白堤，苏堤，杨公堤
湖泊长度	3.2千米（南北）	环湖名街	北山街，南山路，湖滨路
湖泊宽度	2.8千米（东西）	建议游玩时长	3-6小时
		适宜游玩季节	3-5月、9-11月

图2.2.5　百度百科中“西湖”概念条目的信息框

一旦得到了图2.2.5所示西湖概念信息框中信息，就可以将其作为西湖概念的属性描述，在知识图谱中搭建西湖这一概念节点，再将这一概念节点与其他关联的概念节点互联。由于百科中信息框中的信息是由用户输入的，因此通过这一方法构建得到的知识图谱质量较高，但更新较慢。

为了克服这一不足，一些机器学习方法（如聚类和分类等）被提出，这些方法直接从数据中挖掘析取知识。但是，由于学习算法极易受噪声影响以及难以采集足够训练数据来作为训练模型，目前通过机器学习方法来提取海量高质量知识还存在巨大挑战。

为了弥补单一类型文本来构建知识图谱的不足，目前也出现了结合图像、文本等不同类型数据构建的知识图谱，如ImageNet和Vispedia等。

4. 知识图谱的应用

一旦建立起知识图谱，就可基于知识图谱进行知识问答。图2.2.6给出了知识问答过程的示意图。

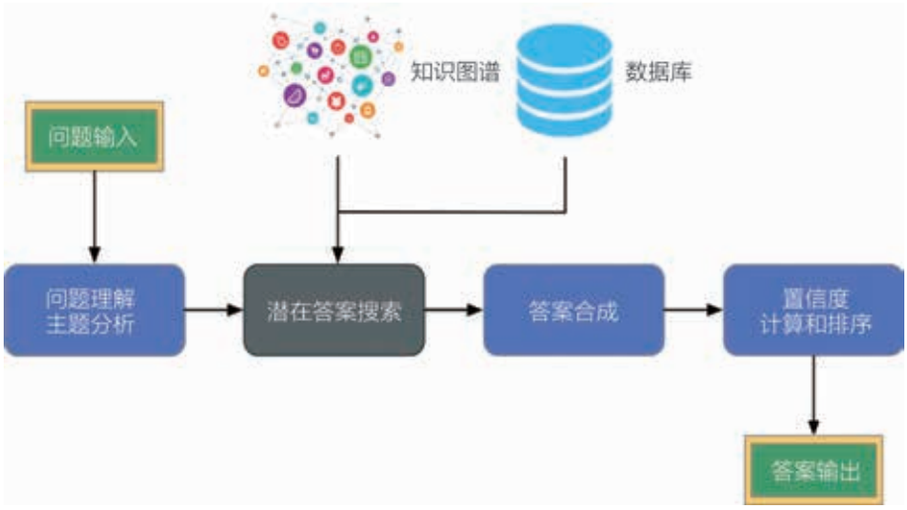


图2.2.6　知识问答（Question-Answer，简称QA）过程示意图

用户给出一个问题后，智能系统首先对问题理解，分析出问题中人名、地名、机构和事件等主题词。然后将这些主题词作为检索请求，向知识图谱和数据库进行检索。数据库里存储着网页和论文等信息。智能系统通过推理，得到潜在答案，并进行答案合成。最后，计算每个答案的置信度（即用户输入问题与潜在答案之间的匹配程度），根据置信度对答案进行排序，将答案输出。

例3　知识图谱推理

如何从知识图谱提供的信息出发推理得到新的知识，是知识图谱推理要解决的问题。

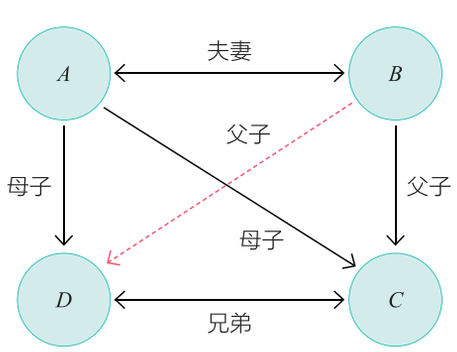


图2.2.7　一个简单的家庭知识图谱

图2.2.7是一个刻画家庭成员关系的简单知识图谱。在这个知识图谱中，节点表示家庭具体成员，边表示家庭成员之间的关系。于是，任意两个具有连接关系的节点可表示一个知识，如节点A，B及连接两者的边线可用“夫妻(A,B)”来表示，说明A和B之间具有夫妻关系。

在上述知识图谱中，如果知道“夫妻(A,B)”和“母子(A,D)”两个知识，那么就可推理出“父子(B,D)”这一个新知识，这就是知识图谱推理。

“父子(B,D)”这一个知识之所以成立，在于我们知道如下推理规则（或者这一条推理

规则可由人类事先输入机器）：如果X和Y具有夫妻关系且X和Z具有母子关系，那么Y和Z具有父子关系，即“夫妻$(X,Y)\wedge$母子$(X,Z)\rightarrow$父子(Y,Z)”。

根据这条规则，只要将X，Y和Z的值分别赋为A，B和D，就可以推理得知“父子(B,D)”成立。

上述过程显示，只要有合适的推理规则，就可从知识图谱中推理得到新的知识（这些知识在知识图谱中不显性表达），从而源源不断地扩充知识图谱。

2.2.6 常识推理

人类习得知识，并非仅仅依靠从规则中推理，或者从海量数据中获得。人类在观察世界和与外界互动中不断积累常识性知识（commonsense knowledge），常识性知识是人类通过归纳、演绎、直觉等综合能力不断积累知识的基础。

虽然当前人工智能已经取得了巨大进展，但是智能算法仍然无法掌握少年儿童所具备的常识。智能算法可模仿人来识别物体、搬运货物、提供在线商务助理等，但仍无法在新情况或非完全信息下采取合适的行动。

与人工智能算法不同，人脑在常识推理方面具有很大潜力，比如：小孩子在看过若干张飞机图像后就能识别出所有飞机；刮风下雨时走在街上，看到前方有块广告牌在风雨中摇晃，我们会避让并向城市管理部门报告（否则广告牌跌落下来会砸伤行人）；见一叶落，而知岁之将暮；审堂下之阴，而知日月之行，阴阳之变；见瓶水之冰，而知天下之寒，鱼鳖之藏也。

上述顿悟、联想和直觉依赖于常识性知识，难以让智能算法从数据中学习。

为了支持常识推理，道格拉斯·莱纳特在1984年提出了Cyc项目，其目标是将人类所有常识编码成机器可读可用的数字化形式，从而可以进行常识推理。截至2015年11月，Cyc包括了23万多个概念、实体和200多万个三元组，但是，Cyc仍难以被有效使用，这也说明数字化人类所有知识的艰辛。Cyc这三个字母取自英文单词“cyclopedia”（百科全书）。

图2.2.8　规则推导智能—人类智能—数据驱动智能

图2.2.8给出了规则推导智能、人类智能和数据驱动智能的不同。由于人类具有学会学习的能力（learning to learn），从而不断积累常识，能够在面对新环境、新问题和新挑战时进行判断、决策和行动。

对于“鸡兔同笼”这样的问题，如果人工智能算法不知晓“鸡和兔各自有多少条腿”这样的常识，仅仅从题目中陈述的内容来解题，那么人工智能算法就无能为力了。

拓展链接

对推理严格约束过程的松绑

“眉头一皱，计上心来”这一现象说明了思维活动具有顿悟形式的非逻辑性特点。一般人工智能需要通过自然语言范畴的谓词、命题和规则等方法在充分定义（well-defined）的前提下进行推理。但是，人脑会有效利用视觉、听觉等难以通过自然语言描述的非严密知识进行推导。因此，人工智能需要模拟大脑这种非严密推导功能，对推理的严格约束进行松绑。为此，钱学森提出了形象思维，潘云鹤提出了综合推理和跨媒体计算，李未提出了开放逻辑等。对推理的松绑才能使推理逐步走向对思维的广泛模拟，这是包含常识推理在内的推理方法值得探究的发展方向。

实践与体验

构建知识图谱

《屈原贾生列传》出自西汉史学家司马迁所创作的《史记》。该篇是屈原和贾谊两人的传记，两人虽不是同一朝代人，但是人生际遇有很多相同之处，如才高气盛、因忠被贬、文学成就卓著等。

阅读下面的《屈原贾生列传》片段，提取人名、地名、书名等概念，并考虑概念的属性，挖掘概念和概念之间的关系，构建与这段文字相关的知识图谱。

“屈原者，名平，楚之同姓也。为楚怀王左徒。博闻强志，明于治乱，娴于辞令。入则与王图议国事，以出号令；出则接遇宾客，应对诸侯。王甚任之……贾生名谊，洛阳人也。年十八，以能诵诗属书闻于郡中。吴廷尉为河南守，闻其秀才，召置门下，甚幸爱。孝文皇帝初立，闻河南守吴公治平为天下第一，故与李斯同邑而常学事焉，乃征为廷尉。廷尉乃言贾生年少，颇通诸子百家之书。文帝召以为博士……太史公曰：余读离骚、天问、招魂、哀郢，悲其志。适长沙，观屈原所自沉渊，未尝不垂涕，想见其为人。及见贾生吊之，又怪屈原以彼其材，游诸侯，何国不容，而自令若是。读鹏鸟赋，同死生，轻去就，又爽然自失矣。”

实践内容：

从《屈原贾生列传》中提取概念、概念属性及概念间的关系，构建知识图谱。将所构建的知识图谱与百度百科相关词条相关联，拓展知识图谱，并可视化

所构建和拓展的知识图谱。

实践步骤：

1. 理解文章所述内容，找出人名、地名、书名、官职名等概念。

2. 定义所找出概念的属性，并给这些属性赋值。

3. 定义概念之间的关系。

4. 将概念与该概念的百度百科词条信息框中的内容关联，拓展所构造的知识图谱。

5. 从知识图谱中挖掘屈原和贾谊之间的关联。

6. 呈现所构造的知识图谱。画出知识图谱，讲解知识图谱中的主要内容。

结果呈现：

将所构造的知识图谱以图2.2.4的形式表现出来。

2.3 基于搜索的问题求解

现实世界中许多问题都可以通过搜索的方法来求解，例如设计最佳的出行线路或是制订合理的课程表。当给定一个待求解问题后，搜索算法会按照事先设定的逻辑来自动寻找符合求解问题的答案，因此一般可将搜索算法称为问题求解智能体。

2.3.1 搜索算法基本概念

广义上的搜索是以探索为手段去寻找问题解决方案。实际上，由于求解的问题具有不同的特点以及算法在求解问题中可以利用不同背景的知识，不同的搜索算法在具体实现过程中会存在较大差异。本节主要针对某些特定类型问题求解的搜索算法进行介绍，为了直观感受所讨论的问题类型，先看下面的例子。

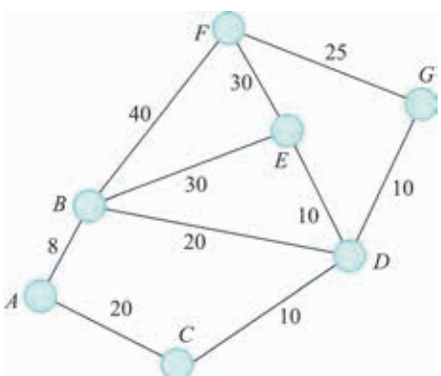


图2.3.1 某市公交线路示意图（单位：千米）

例如，小华不久前搬到某座城市，对当地的交通还不熟悉。某天，小华决定去拜访该城市中的一名朋友，正当他向朋友询问其具体地址时，朋友发来了一张公交线路示意图（如图2.3.1所示），并声称它能指引小华到达目的地。

看过地图后，小华知道图中标识为A，B，C，D，E，F和G的节点对应着几个重要的公交换乘站。任意两个公交换乘站之间如果有连线，表示它们之间存在可以往返的公交线路，连线上的数字表示所相连两个公交换乘站之间的距离。在图2.3.1中，公交换乘站A是距离小华家最近的一个车站，G是小华要去的目的地。小华现在需要找到一条从A到G的路线，并且希望找到的路线所花费的时间最少。由图2.3.1可知，从A到G距离最短的出行路线是$A\rightarrow B\rightarrow D\rightarrow G$，其距离是38千米。那么，怎样设计一个算法从一张地图中寻找任意两个站点之间的最短路径呢？

拓展链接

中国邮递员问题

中国邮递员问题（Chinese postman problem）由我国数学家管梅谷于1960年首次提出，这是一个著名的图论问题，引起了世界上不少数学家的关注。这个问题可做如下描述：一名邮递员从邮局出发投递邮件，然后返回邮局，途中他必须经过由他负责投递的每条街道至少一次。如何为这名邮递员设计一条投递路线，使其耗时最少？

计算机并不明白人类为何能够“一眼”找出一条从A到G的最短路线，计算机程序能够做到的是按照既定规则（如在当前节点寻找一个最优后续节点），从初始点出发，不断尝试从一个节点移动到下一个节点，直到算法到达目标节点。在图2.3.1中，与初始点A相邻的后续节点只有B和C，因此要从初始点A出发，找到一条到达终点G的路线，算法必然要从A节点移动到B或C节点。接着算法从选中的B或C节点出发，继续选定其后续节点，如此不断尝试，直到到达终点G为止。

在详细描述搜索算法之前，先介绍若干概念。

- 状态。状态可以认为是搜索算法在某一时刻所处的位置，相应地，搜索算法在开始和结束时所处的位置称为初始状态和终止状态。如在图2.3.1中，搜索算法的初始状态位于站点A，终止状态位于站点G。
- 动作。动作指的是搜索算法从一个状态转变到另外一个状态所采取的行为。一般假设在每个状态下所能够采取的行为数量都是有限的。如在图2.3.1中，搜索算法可采取行动从状态A到达状态B，像这样采取行动从一个状态到另一个状态的过程叫作状态转移。
- 路径。完成一系列连续的状态转移所得到的状态序列就构成了从起点到终点的路径。如从状态A到状态B，接着到状态D，最后到状态G，就形成了$A\rightarrow B\rightarrow D\rightarrow G$这样的一条路径。很显然，沿着这条路径出行所花费的代价是路途中乘车所花费的总时间。在路径搜索问题中，任何一条路径的代价都不会是负数。
- 测试目标。用于判断当前状态是不是目标状态。在交通问题中，小华的目的地是G，因此目标测试只需判断当前状态是否为G即可。当然，即使到达了目标状态，找到的路径也未必是代价最小的。

2.3.2 搜索算法

2.3.1小节给出了搜索算法的一个大致轮廓，即搜索算法不断从某一状态转移到下一状态，直到到达终止状态为止。

在搜索算法中，从当前状态出发寻找后续节点，一般会面临多种选择。例如在图2.3.1中，从A出发，可以选择B或C；从B出发，可以选择A，D，E或F；从C出发，可以选择A或D。可见，状态之间的这种转移构成了如图2.3.2所示的分层树状结构，该结构称为搜索树。

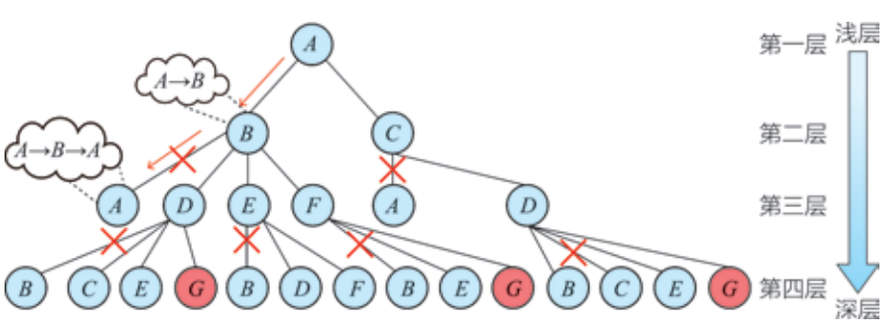


图2.3.2　交通问题对应的搜索树的一部分

在搜索树中，每个节点可用一个状态来标记，表示从根节点出发，经过怎样的路径到达该节点，两个节点之间的连线表示这两个节点之间存在状态转移。在搜索树中，由于每个节点用状态来标记，因此可能存在两个拥有相同标记的节点，但其含义不同。如在图2.3.2中，根节点A（即第一层的唯一节点）和第三层左侧第一个节点A虽然符号相同，但其含义不同：根节点A表示搜索算法位于初始节点状态、正准备向后续节点移动，第三层左侧第一个节点A表示从初始节点A出发、经过节点B、回到节点A这条路线所表示的状态。

也就是说，搜索树中每个节点刻画了从初始节点（根节点）到该节点的一条唯一的路径，这条唯一的路径将搜索树中的节点彼此区分开来。如在图2.3.2中，第二层最左侧标记为“B”的节点对应了一条路径$A\rightarrow B$；同理，第三层最左侧标记为“A”的节点对应了一条路径$A\rightarrow B\rightarrow A$。

为了简化表达，本节对节点和状态这两个概念不做区分，只在必要情况下做说明。

需要注意的是，在路径搜索过程中不能出现回路。如在图2.3.1中，假设当前已经找到的路径是$A\rightarrow B\rightarrow E\rightarrow D$，即搜索算法当前位于状态$D$。从状态$D$出发，可以选择$B$，$C$或$G$。显然，选择$B$将产生回路（即一条从$B$出发回到$B$的路径），这样一方面造成了无意义地把时间浪费在回到原地的路径上，另一方面还产生了沿着环路不停绕圈的死循环问题。因此，如果某个动作会使状态转移到一个已被访问的节点，那么搜索算法将不会采取这个动作。

如图2.3.2所示，搜索算法会在“×”所示位置将存在回路（即存在无限长度的路径）的分支切断，余下的部分才是搜索算法实际上要构建的搜索树。

这样，搜索算法从初始节点（对应初始状态）出发，不断选择后续节点，完成了搜索树的构造。一开始，搜索树中只有根节点。在每一步中，搜索算法将选择与搜索树中某个节点相邻的一个后续节点加入搜索树，这个操作叫作扩展一个节点。不难发现能够被扩展的节点必须满足两个条件：（1）该节点不能已经在搜索树中，即该节点不能已经被扩展过；（2）该节点能够从搜索树中某个节点出发通过执行一个动作直接到达，即被扩展节点和搜索树中的某个节点是相邻的。这些能被扩展的节点构成的集合称为未访问节点集合。

于是，搜索算法的每步操作可以做如下描述：每次选择未访问节点集合中的一个节点加入当前搜索树，检查这个节点的所有后续相邻节点，将满足条件的节点加入未访问节点集合中，重复执行上述操作直至被扩展的节点对应一条从初始节点到终止节点的路径。

例如在图2.3.3中，实心节点A和B是已经加入搜索树的节点，从A能够直接到达的节点有B和C，从B能够直接到达的节点有A，D，E和F。由于路径$A\rightarrow B$已经在搜索树中，因此B不能再次被扩展。而路径$A\rightarrow B\rightarrow A$产生了回路，也应当舍去。所以此时的未访问节点集合是$\{C, D, E, F\}$。

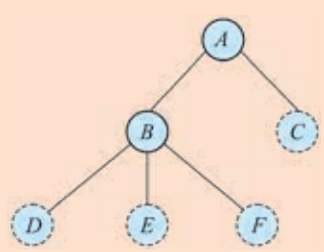


图2.3.3　未访问节点示意图

注：图中实线节点表示当前搜索树中已经被访问过的节点，虚线节点为当前未访问节点集合中的节点。注意图中已经去掉了产生回路的分支。

2.3.3 深度优先搜索和广度优先搜索

一般未访问节点集合中会包含若干节点，那么搜索算法应该按照怎样的原则从该集合中取出一个节点？在实际应用中，最常见的两种从未访问节点集合中取出节点的策略是深度优先搜索和广度优先搜索。

深度优先搜索总是沿着某个分支进行搜索，直至不能再深入为止，即优先扩展搜索树当前未访问节点集合中最深的节点。深度优先搜索算法在搜索过程中总是倾向于沿着一个分支前进，直到该分支上所有节点都被访问完，再返回上一层进行另一轮深度优先搜索。

例如在图2.3.4中，搜索算法从初始节点A出发可选择后续节点B或C，假设算法这时选择B作为后续节点，而将节点C放入未访问节点集合。在到达B以后，算法需要从D，E，F中选择一个作为后续节点，假设选择节点D，接着到达节点D后选择C作为后续节点。这时，算法发现在节点C已无任何节点可以作为后续节点，即从C出发已经没有合适的动作可以选择。于是，算法只好返回节点D，将节点E作为后续节点，继续向前搜索。

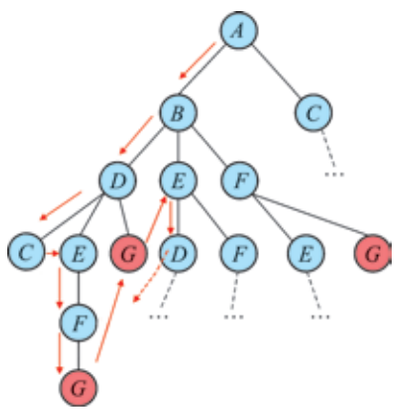


图2.3.4 深度优先搜索的扩展顺序

广度优先搜索总是优先扩展未访问节点集合中最浅的节点，其过程如图2.3.5所示。广度优先搜索在执行中倾向于优先把同一层的所有可能节点访问完后，再考虑进行更深层的探索。

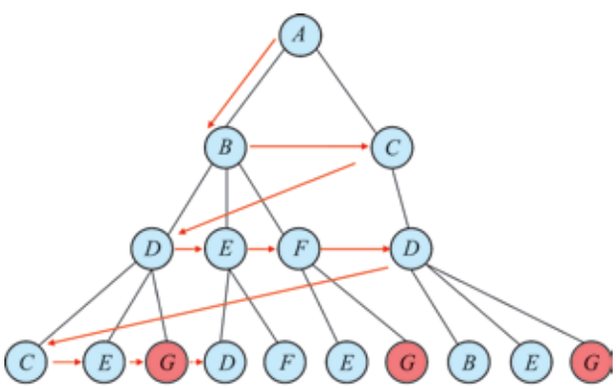


图2.3.5 广度优先搜索的扩展顺序

例如在图2.3.5中，一开始搜索算法在节点A（其后续节点是B和C），尝试扩展节点B而把节点C加入未访问节点集合。这时，算法发现节点B有D，E，F三个后续节点，于是将这三个节点加入未访问节点集合，但算法并没有对节点B的这三个后续节点进行扩展，而是选中了更浅的节点C来进行下一步的探索。这样，算法在扩展第三层节点之前，先完成了对第二层所有节点的探索。

需要强调的是，对于一个搜索问题，只要存在答案（即从初始节点到终止节点存在满足条件的一条路径），那么排除了回路的深度优先搜索和广度优先搜索均能找到一个答案，但这个答案不一定是最优的（如距离最短）。

拓展链接

搜索算法的实现

在实现深度优先搜索和广度优先搜索时，通常会使用递归的技巧和队列的数据结构。递归是一种重要的编程技术，用于让一个函数从其内部调用其自身。当一个问题依赖于若干个形式相同但更加简单的子问题时，就可以用递归的方法求解，在编写程序时递归通常体现为在一个函数中调用其自身。队列保存了一组数据，可向其中加入数据和取出数据。队列保证了每次被取出的数据一定是当前保存的数据中最早被加入的。

2.3.4 启发式搜索

在以上几个小节探讨的算法中，搜索算法并没有考虑问题本身之外的其他信息。在有些情况下，搜索算法可以利用的信息不仅包括问题中的定义（如公交站点之间的连线），还包括一些与这个问题相关的辅助信息，这些辅助信息是否能够被用来提升搜索算法的效率呢？

这是有可能的，比如在上述的交通路线问题中，假设事先知道了从初始节点到终止节点的最短路线不会通过公交车站 C，那么在搜索过程中就可以忽略所有到节点 C 的状态转移，这样可以显著减小搜索树的规模，从而提升搜索算法的效率。像这样在搜索过程中利用问题定义以外辅助信息的搜索算法，称为启发式搜索算法，或者叫有信息的搜索算法。

在路径搜索问题中，可引入任意一个站点与目标站点之间的直线距离作为辅助信息，来提升搜索算法的效率。这个辅助信息对最短路径搜索具有很大的帮助，因为与目标站点直线距离越近的车站，其到达目标站点所需的时间越短，也就越有可能出现在最短路径中。

根据这一想法，可以设计一个直观的最短路径搜索算法：算法从初始节点开始，每一步都将未访问节点集合中离目标站点直线距离最近的节点加入搜索树，直至到达目标站点。这个算法称为贪婪最佳优先搜索。

表 2.3.1　辅助信息：交通路线问题中各个站点到目标站点 G 的直线距离

站点	A	B	C	D	E	F	G
距离	30	20	19	10	5	25	0

表 2.3.1 中定义了每个站点到目标站点 G 的直线距离。在给定表 2.3.1 的基础上，贪婪最佳优先搜索算法的流程如图 2.3.6 所示。

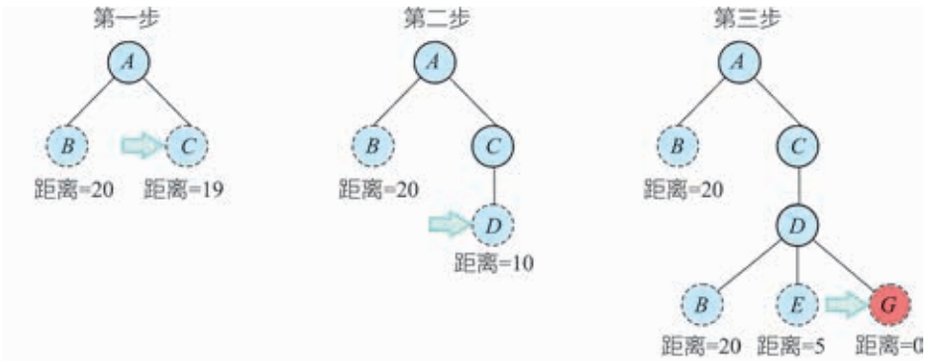


图2.3.6　贪婪最佳优先搜索算法的搜索过程

注：节点下给出了当前状态所在站点到目标站点的直线距离。

搜索算法从站点A出发时，可以选择站点B或C作为后续节点。由于站点C到目标站点G的直线距离比站点B到目标站点G的直线距离更近，因此算法“贪婪地”将站点C选为A的后续节点。在选择站点C后，由于站点C仅有一个后续节点D，因此算法选择站点D作为后续站点。在节点D时，由于站点D和目标站点G是相邻节点，算法直接将节点G选为后续节点，于是到达目标站点G，完成搜索任务。这样，从初始节点A出发，贪婪最佳优先搜索算法在额外信息的帮助下，找到一条从A到G的路线$A\rightarrow C\rightarrow D\rightarrow G$。

但是，在“贪婪”机制下找到的路径$A\rightarrow C\rightarrow D\rightarrow G$并非是最短路径，因为$A\rightarrow B\rightarrow D\rightarrow G$的路径开销比$A\rightarrow C\rightarrow D\rightarrow G$的路径开销要小。产生这样的搜索结果，其原因是：贪婪最佳优先搜索算法在当前节点时，每次均“贪婪地”从与当前节点相邻的节点中选择如下节点作为后续节点：与目标节点直线距离最近的节点。这样就会造成贪婪最佳优先搜索算法过于重视当前最优，而忽视了全局最优。

为了克服这一不足，另外一种启发式搜索算法——A*算法被提出。

从上面的内容可知，搜索算法每次都是从未访问节点集合中选择一个“最佳”节点，作为后续节点。于是，可以定义一个关于节点的评价函数$f(n)$。评价函数$f(n)$给出一个对节点n的评分值，于是算法可根据$f(n)$的大小来判断是否优先扩展节点n、将其作为后续节点。这样，每次进行节点扩展时，搜索算法计算未访问节点集合中每个节点评价函数的大小，选取评价函数值最小的节点来扩展。

很显然，在贪婪最佳优先搜索中，评价函数只计算节点与目标节点的直线距离，优先扩展与目标节点直线距离最短的节点。

那么A*算法是如何克服贪婪最佳优先搜索算法的不足，保证找到的路线一定是最短的呢？不妨先思考一下贪婪最佳优先搜索算法的不足。根据图2.3.6中的流程，贪婪最佳优先搜索算法第一步就偏离了实际的最短路线，“贪婪地”选择了节点C，原因在于其认为节点C与目标节点之间的直线距离更近（相对于节点B而言）。为什么节点C不是一个更好的后续节点呢？这是因为从节点A到节点C的代价比从节点A到节点B的代价大，而贪婪最佳优先搜索算法并没有将这个因素考虑在内。

从这个角度出发，在从初始节点逐步向前搜索的过程中，为了保证搜索得到的一条

从初始节点到目标节点的路径是最短的，则可将初始节点到目标节点的距离分成两部分：（1）初始节点到当前节点的路径代价；（2）当前节点到目标节点之间的直线距离。将两者之和作为评价函数的取值大小。根据这样定义的评价函数来选择后续扩展的节点，这就是A*算法的思路。

具体而言，对于未访问节点集合中的某个节点n，A*算法评价节点n取值大小的评价函数$f(n)$由两部分构成：

函数$g(n)$：表示从初始节点到当前节点n的实际距离。

函数$h(n)$：表示当前节点n到目标节点的直线距离。函数$h(n)$也称为启发函数。

于是，A*算法的评价函数定义为$f(n)=g(n)+h(n)$。在这种评价函数评估中，只有满足以下条件的节点才优先被扩展为后续节点：从初始节点到该节点的实际距离与该节点到目标节点的直线距离之和最小。

对于如图2.3.1所示的最短路径搜索问题，A*算法的搜索过程如图2.3.7所示，每个未被访问节点下方列出的算式中三个数字从左到右分别对应该节点$g(n)$，$h(n)$和$f(n)$的取值大小。可见A*算法每次选择$f(n)$值最小的节点作为后续节点。

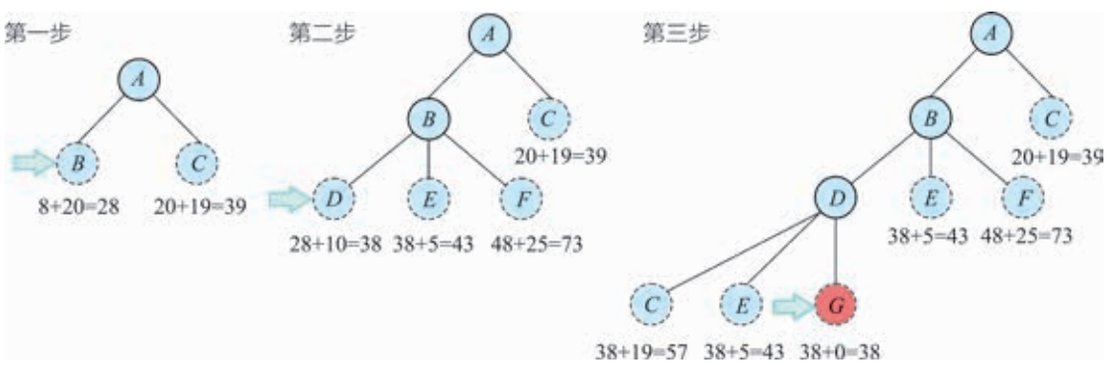


图2.3.7　最短路径搜索问题中A*算法的搜索过程

如图2.3.7所示，搜索算法第一步从初始节点A出发，节点A可选择的后续节点是B或C，其中从A到B的实际距离是8，B到目标节点的直线距离是20，两者相加等于28，这比A到C的实际距离加上C到目标节点的直线距离之和39要小，因此搜索算法根据评价函数取值大小将节点B作为后续节点，同时将D，E，F加入未访问节点集合。第二步分别计算未访问节点集合$\{C, D, E, F\}$中每个节点的评价函数取值，算法发现$f(D)$取值最小，因此将节点D作为后续节点，并将$D\rightarrow C$、$D\rightarrow E$和G加入未访问节点集合。注意：$D\rightarrow C$和$D\rightarrow E$两个节点分别对应的路径是$A\rightarrow B\rightarrow D\rightarrow C$和$A\rightarrow B\rightarrow D\rightarrow E$，它们不同于前面已经加入未访问节点集合的$C$和$E$（这里的$C$和$E$两个节点分别通过路径$A\rightarrow C$和$A\rightarrow B\rightarrow E$抵达）。为了避免歧义，文中在它们的左下角标出了其前驱节点。在第三步时，搜索算法找到了最短路径$A\rightarrow B\rightarrow D\rightarrow G$。

与贪婪最佳优先算法不一定能够找到最短路径不同，A*算法找到的路径一定是最短路径。另一方面，由于A*算法能够利用辅助信息，因此它比其他算法用更少的步骤。

在实际中，A*算法的性能表现取决于启发函数的设计，只要定义一个合适的启发函数，A*算法就能大幅缩短搜索所需的时间。

思考与练习

1. 以下算法流程总结了本小节中介绍的搜索算法。在深度优先搜索、广度优先搜索和A*算法中，横线（1）和（2）处分别应该填写什么内容？

```
未访问节点集合←{初始节点}
若未访问节点集合非空，则循环执行以下步骤：
    对未访问的每个节点，计算_____（1）_____
    节点n←_____（2）_____
    未访问节点集合←未访问节点集合−{n}
    若n是目标节点，则记录路径并退出循环
    未访问节点集合←未访问节点集合∪{从n可转移而抵达的不产生环路的节点}
```

2. 假设图2.3.8是与图2.3.1类似的一张线路示意图。如果使用深度优先搜索求状态0到状态4的一条路径，我们可以用下表来模拟搜索过程。注意：在下表中，节点的深度定义为它对应路径中状态转移的次数，如果多个未访问节点的深度相同，那么在这个例子里算法优先选择状态编号大的节点。

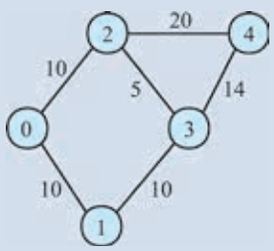


图2.3.8　线路示意图

步骤	当前状态	当前未访问节点集合（用上划线标出了下一个扩展的节点）
1	0	深度1：{0→1,$\overline{0\rightarrow 2}$}
2	2	深度1：{0→1} 深度2：_____（1）_____
3	_____（2）_____	找到路径0→2→4

请仔细观察上表中各项内容的含义，根据深度优先搜索的思路，在横线（1）和（2）处填写内容。问：找到的路径0→2→4是代价最小的吗？

3. 假设图2.3.8中各个状态到状态4的直线距离如表2.3.2所示，请模仿课文中的做法，尝试使用贪婪最佳优先算法求出一条从0到4的路径。问：求出的路径是不是代价最小的路径？

表2.3.2　状态间的直线距离

状态	0	1	2	3	4
距离	22	23	20	14	0

4. 根据表2.3.2，试用A*算法求出状态0到状态4的最优路线。

2.4 决策树

决策树是一种通过树形结构进行分类的方法。在决策树中，树形结构中每个节点表示对分类目标在属性上的一个判断，每个分支代表基于该属性做出的一个判断，最后树形结构中每个叶子节点代表一种分类结果。

2.4.1 决策树分类概念

决策树将分类问题分解为若干基于单个信息的推理任务，采用树状结构来逐步完成决策判断。事实上，人们在逻辑推理过程中经常使用决策树的思想。

下面通过一个例子来解释决策树的分类。一个游乐场经营者希望通过所收集的一系列气象参数来预测游客是否会来游乐场游玩。为了实现这样的预测，游乐场经营者希望通过天气状况（如晴、雨、多云）、温度高低、湿度大小、风力强弱四个气象特点来预测游客是否来游乐场游玩。

通过长时间的观察，游乐场经营者得到如表2.4.1所示数据，记录了在不同天气、温度、湿度和风力情况下，游客是否来游乐场的情况。

表 2.4.1　气象特点与游客是否来游乐场的关系

序号	天气	温度（℃）	湿度	是否有风	是(1)否(0)前往游乐场
1	晴	29	85	否	0
2	晴	26	88	是	0
3	多云	28	78	否	1
4	雨	21	96	否	1
5	雨	20	80	否	1
6	雨	18	70	是	0
7	多云	18	65	是	1
8	晴	22	90	否	0
9	晴	21	68	否	1
10	雨	24	80	否	1
11	晴	24	63	是	1
12	多云	22	90	是	1
13	多云	27	75	否	1
14	雨	21	80	是	0

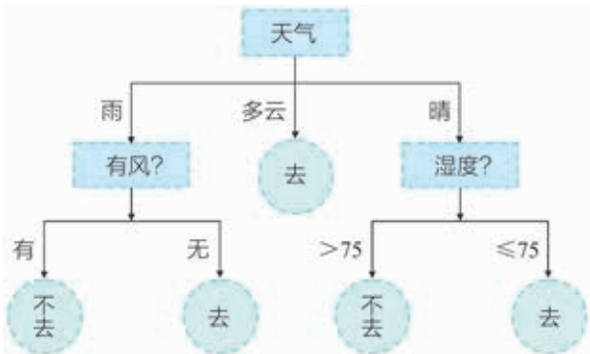


图2.4.1　游乐场游玩问题决策树

观察表2.4.1，可以画出如图2.4.1所示的决策树。由图可知：

- 第一层是天气状况，具有雨、多云和晴三种属性取值。
- 当天气状况的属性取值为“多云”时，样本子集为{3, 7, 12, 13}，“多云”属性可将样本子集{3, 7, 12, 13}划分为“前往游乐场游玩”一个类别。
- 当天气状态的属性取值为“晴”时，样本子集为{1, 2, 8, 9, 11}，可知天气状况为“晴”这个属性所得到的样本子集{1, 2, 8, 9, 11}不属于同一个类别，要进一步使用其他属性对这个样本子集进行划分。经观察，通过“湿度是否>75？”这一属性值，可以将该样本子集进一步划分成{1, 2, 8}（>75，不前往游乐场）和{9, 11}（≤75，前往游乐场）两个样本子集。此时这两个样本子集已被划分为两类，不需要再划分。
- 当天气状况的属性取值为“雨”时，样本子集为{4, 5, 6, 10, 14}，此时样本子集无法被划分为一类，需要进一步使用其他属性对这个样本子集进行划分。经观察，通过“有风否？”这个属性值，可将该样本子集进一步划分成{4, 5, 10}（无风，前往游乐场）和{6, 14}（有风，不前往游乐场）两个样本子集。此时这两个样本子集已被划分为两类，不需要再划分。

在这里可以看出，“温度”这一特点及其属性值在决策树构造过程中没有使用。对于将表2.4.1所记录样本进行分类这个问题而言，“温度”是多余的特点。

实际上整个建立决策树的过程，就是选择一个属性值，基于这个属性对样本集进行划分，得到子集划分结果。再选择其他属性，对得到的划分结果进行划分，直至最后所得划分结果中每个样本为同一个类别。在表2.4.1中，就是要将所有样本通过不同属性的层层划分，分为两个类别（即“去游乐场”和“不去游乐场”）。

如果需要判断某一天游客是否会来游乐场游玩，只要了解当天的天气、温度、湿度和风力等属性，根据已建立的决策树，层层递进推断即可。

使用决策树进行分类的优点在于其决策过程由树形结构来完成，易于理解，比如可清楚地了解在决策的每一步依次使用了何种属性来进行判断。

2.4.2 构建决策树

构建决策树来解决实际生活中的问题时，需按照一定的顺序选择划分属性。通常，性能好的决策树随着划分不断进行，决策树分支节点的“纯度”会越来越高，即其所包含样本尽可能属于相同类别。

为了逐次选出最优属性，可以采用信息增益（information gain）这一指标。信息增益被用来衡量样本集合复杂度（不确定性）所减少的程度。

在社会生活中，我们不断从外界接收信息。对于不同的信息，如何去判断信息量的大小呢？比如，“太阳明天从东边升起”和“明天可能会遇见月全食”这两条信息，很显然两者蕴含的信息量不一样。前者是确定性信息，信息量小；后者具有不确定性，信息量较大。

香农提出了“信息熵”（entropy）的概念，用来度量信息量的大小。直觉上而言，对一个不确定的信息（很少出现）的了解需要更多的信息。相反，如果对某件事已经有了较多了解，就不需要太多信息了。从信息论的角度来看，对信息的度量等于计算信息不确定性的多少。

拓展链接

香农与信息熵

信息熵是信息论中最基本、最重要的一个概念，由香农于1948年提出，其表达了信息的“不确定性”（uncertainty）。

熵（entropie，英文为entropy）这一概念最早在热力学领域由德国物理学家克劳修斯于1854年提出。汉语中本无“熵”字，我国物理学家胡刚复教授于1923年首次把entropie译为“熵”。熵在热力学中用来衡量系统的混乱程度，一个系统的混乱度越高，它的熵就越高。

香农将熵这个概念从热力学领域应用于信息领域，其提出的“信息熵”或“香农熵”解决了信息的量化度量问题。一条信息所承载的信息量大小和它的不确定性有直接关系。比如，球队甲和球队乙在历史上进行过多次对决比赛，甲队战胜乙队的胜率为x，乙队战胜甲队的胜率为$1-x$。甲、乙两队历史上胜负信息的信息熵为$-[x\log_2 x+(1-x)\log_2(1-x)]$。在$x$取值为$\frac{1}{2}$时，信息熵取得最大值1。也就是说甲、乙两队势均力敌时，两队对决的胜负不确定性最大，信息熵取值最大。

“假设有K个信息，其组成了集合样本D，记第k个信息发生的概率为$p_k(1\leqslant k\leqslant K)$”。如下定义这$K$个信息的信息熵：

$$E(D)=-\sum_{k=1}^{K} p_k \log_2 p_k$$

$E(D)$值越小，表示D包含的信息越确定，也称D的纯度越高。需要指出，所有p_k累加起来的和为1。

现在应用熵这个度量标准来构建决策树。表2.4.1中14个样本分属于“游客来游乐场”（9个样本）和“游客不来游乐场”（5个样本）两个类别，即K=2。

记“游客来游乐场”和“游客不来游乐场”的概率分别为p_1和p_2，显然$p_1=\frac{9}{14}$，$p_2=\frac{5}{14}$，则这14个样本所蕴含的信息量（熵）如下计算：

$$E(D)=-\sum_{k=1}^{2}p_k\log_2 p_k=-\left(\frac{9}{14}\times\log_2\frac{9}{14}+\frac{5}{14}\times\log_2\frac{5}{14}\right)=0.940$$

表2.4.1中有天气状况、温度高低、湿度大小、风力强弱四个气象特点，下面计算这四个特点所对应的信息熵。

以天气状况为例，天气状况包含“雨”“晴”“多云”三个属性。这三个属性对14个样本进行划分，在决策树中产生了三个分支节点，每个分支节点所得到的样本子集就是其对应属性对数据的划分。例如，“多云”这一属性包含四个样本{3, 7,12,13}。

当计算天气状况所包含“雨”“晴”“多云”三个属性信息增益时，其取值情况和对应样本集情况如表2.4.2所示。这里记属性取值为a_i，如天气状况的三个属性记为a_0=“晴”，a_1=“多云”，a_2=“雨”，记属性可能取值总数为n，那么用这些属性划分样本集可得到n个分支节点。属性取值为a_i，对应分支节点所包含子样本集记为D_i，该子样本集包含样本数量记为$|D_i|$。

表 2.4.2　天气属性样本情况统计

天气属性取值a_i	“晴”	“多云”	“雨”
对应样本数$\|D_i\|$	5	4	5
正负样本数量	（2+, 3−）	（4+, 0−）	（3+, 2−）

根据表2.4.2的统计情况，可如下计算天气状况每个属性值（即每个分支节点）的信息熵：

$$\text{“晴”：}\ E(D_0)=-\left(\frac{2}{5}\times\log_2\frac{2}{5}+\frac{3}{5}\times\log_2\frac{3}{5}\right)=0.971$$

$$\text{“多云”：}\ E(D_1)=-\left(\frac{4}{4}\times\log_2\frac{4}{4}+0\right)=0$$

$$\text{“雨”：}\ E(D_2)=-\left(\frac{3}{5}\times\log_2\frac{3}{5}+\frac{2}{5}\times\log_2\frac{2}{5}\right)=0.971$$

一旦计算得到天气状况三个属性值（晴，多云，雨）的信息熵后，可进一步计算天气状况的信息增益。

天气状况的信息增益如下计算：

$$Gain(D, A)=E(D)-\sum_{i=1}^{n}\frac{|D_i|}{|D|}E(D_i)$$

其中，A=“天气状况”。于是天气状况这一气象特点的信息增益为：

$$Gain(D,\text{天气})=0.940-\left(\frac{5}{14}\times 0.971+\frac{4}{14}\times 0+\frac{5}{14}\times 0.971\right)=0.246$$

同理，可计算温度高低、湿度大小、风力强弱三个气象特点的信息增益。（注：表2.4.1中湿度大小具有两个属性取值，即大于75为一种属性取值，小于或等于75为另一种属性取值）

通常情况下，某个分支的信息增益越大，则该分支对样本集划分所获得的“纯度”越大、信息不确定性减少的程度越大。

思考与练习

1. 计算表2.4.1中温度高低、湿度大小、风力强弱三个气象特点的信息增益。将天气状况、温度高低、湿度大小、风力强弱作为分支点来构造图2.4.1决策树时，是否信息增益大的气象特点离根节点越近？

2. 在表2.4.1中，天气状况所对应属性“多云”的信息熵计算为0，解释其意义。

3. 在计算某段文字的信息熵时，可通过计算文字中所包含单词的出现概率来实现，与文字所蕴含的内容无关。信息熵的提出背景是为了信息通信，以减少信息冗余为主要目的。一段文字的信息熵小，说明其确定性高，但不能用信息熵来直接刻画文字语义内容的丰富与平凡。如《诗经·郑风》中“山有扶苏，隰有荷华。不见子都，乃见狂且”这句话，内容让人浮想联翩，但信息熵却不一定大，这也说明了“一字抵千金”和“一图胜千言”中信息内容与信息出现频率有时候关联不大的特点。试举其他类似的例子。

4. 如表2.4.3所示，每朵鸢尾花有萼片长度、萼片宽度、花瓣长度、花瓣宽度四个特征。现在需要根据这四个特征将鸢尾花分为杂色鸢尾、维吉尼亚鸢尾和山鸢尾三类，试构造决策树进行分类。

表2.4.3　鸢尾花的相关数据

萼片长度	萼片宽度	花瓣长度	花瓣宽度	种类
5.0	2.0	3.5	1.0	杂色鸢尾
6.0	2.2	5.0	1.5	维吉尼亚鸢尾
6.0	2.2	4.0	1.0	杂色鸢尾
6.2	2.2	4.5	1.5	杂色鸢尾
4.5	2.3	1.3	0.3	山鸢尾

实践与体验

计算文章的信息熵

有些科学家发现，对于所有语言来说，词语排列中信息总量都是相同的，即使这些语言之间毫无联系。这种一致性暗示着所有语言都有个共同的祖先，或者人类大脑在处理语言时存在普遍性。

也有科学家发现，中文信息熵大于英文信息熵，意味着不确定性也大，表明中文因词义丰富，复杂程度高于英文。

收集中英文对照的短文，在计算短文内中文单词和英文单词出现概率的基础上，计算短文信息熵，比较中文短文信息熵和英文短文信息熵的大小。

实践内容：

计算中英文对照两篇短文的信息熵，体会信息熵概念。给定一篇短文，在信息熵计算公式 $E(D)=-\sum_{k=1}^{K} p_k \log_2 p_k$ 中，$p_k(1\leqslant k\leqslant K)$ 表示短文内第 k 个单词出现的概率，K 为该短文中所有不同单词的数目。

实践步骤：

1. 收集新闻类、科技类或法律类的中英文对照短文，找出每篇短文内的单词。
2. 根据信息熵公式计算每篇短文的信息熵。
3. 信息熵反映了所表达信息的不确定性大小。对计算结果进行分析，科技类或法律类短文的信息熵是否小（不确定性小）？

结果呈现：

将计算结果以表格形式呈现，对比不同类型的文档结果。

2.5 回归分析

在现实生活中，往往需要分析若干变量之间的关系，如碳排放量与气候变暖之间的关系、某商品的广告投入量与销售量之间的关系等，这种分析不同变量之间存在关系的研究叫回归分析，刻画不同变量之间关系的模型称为回归模型。一旦确定了回归模型，就可以进行预测等分析工作，如从碳排放量预测气候变化程度、从广告投入量预测商品销售量等。

2.5.1 回归分析的基本概念

表2.5.1给出了莫纳罗亚山（夏威夷岛的活火山）从1970年到2005年间每5年的二氧化碳浓度，单位是百万分比浓度（parts per million，简称ppm）。

表 2.5.1 莫纳罗亚山从1970年到2005年间每5年的二氧化碳浓度

年份x	1970	1975	1980	1985	1990	1995	2000	2005
CO_2 浓度y（ppm）	325.68	331.15	338.69	345.90	354.19	360.88	369.48	379.67

如果从表2.5.1所给出的数据出发，能够得到该地区每5年二氧化碳浓度变化的数学模型（即时间年份和二氧化碳浓度之间的关联关系），由此可以对未来每5年的二氧化碳浓度进行预测。

经初步观察，发现莫纳罗亚山地区的二氧化碳浓度在逐年缓慢增加，因此使用简单的线性模型来刻画时间年份和二氧化碳浓度两者之间的关系，即二氧化碳浓度$=a\times$时间$+b$。设时间年份为x、二氧化碳浓度为y，即$y=ax+b$。表2.5.1提供了8组数据，每一组数据是(x, y)的一个组合。那么，可利用这8组数据来确定模型中参数a和b的值。一旦求解出a和b的值，输入任意的时间年份（甚至是1970年之前的时间年份），该模型可估算出该时间年份所对应的二氧化碳浓度值。这种建立变量之间关联关系，且利用这种关联关系进行预测分析的方法叫回归分析。

上述回归分析过程是一种“有监督学习”（supervised learning），与之相对的学习过程叫“无监督学习”（unsupervised learning）。在有监督学习过程中，需要给出包含所有变量的训练数据。

在回归分析中，刻画数学关系的模型包含了一些未知参数（如$y=ax+b$中的参数a和b），这些参数需要从已有数据中计算得到。那么如何预设一个合理的模型？又如何对模型中的未知参数进行计算呢？

一般情况下，为了简化问题，往往假设模型是符合线性分布的。符合线性分布的模型

结构相对简明，参数容易计算。而不符合线性分布的模型（即非线性模型）一般难以直接定义，通常需要借助一些先验知识来完成。例如，在进行一些复杂的经济模型的预设时，由于线性模型难以刻画变量之间的复杂关联因素，需要使用非线性模型。

拓展链接

回归思想的提出

英国著名生物学家兼统计学家高尔顿（生物学家达尔文的表弟）在研究父母身高和子女身高时发现：相比于父母身高，子女身高有“衰退”（regression）（也称作“回归”）的倾向。假设父母平均身高为x、子女平均身高为y，高尔顿发现x和y之间存在如下线性关系：$y=33.73+0.516x$。可以看到，父母平均身高每增加一个单位，其成年子女平均身高只增加0.516个单位，它反映了一种“衰退”效应（“回归”到正常人平均身高）。虽然后来研究发现x和y之间并不总是具有“衰退”（回归）关系，但是为了纪念高尔顿这位伟大的统计学家，“线性回归”这一名称就保留下来了。

2.5.2 回归分析中参数计算

最简单的线性回归模型就是上述使用的一元线性回归模型，它只包含一个自变量x和一个因变量y，并且假定自变量和因变量之间存在如$y=ax+b$的线性关系。

为了求解参数a和b，需要给定若干组(x, y)数据，然后从这些数据出发来计算参数a和b。图2.5.1给出了莫纳罗亚山从1970年到2005年间每5年的二氧化碳浓度的8组(x, y)数据，一元线性回归分析实际上就是寻找“$ax+b$”形成的一条直线，使得这条直线尽可能靠近或穿过这8组(x, y)数据，即能够以最小的误差来拟合这8组(x, y)数据。

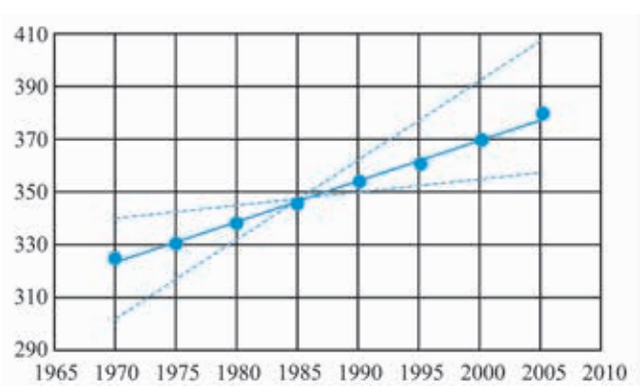


图2.5.1　莫纳罗亚山地区时间年份与二氧化碳浓度之间的一元线性回归模型（实线为最佳回归模型）

在一元线性回归模型中，最关键的问题就是如何计算参数a和b使误差最小化。那么，误差是怎么表示的呢？

在图2.5.1中，横坐标取值为时间年份x，纵坐标取值为二氧化碳浓度y，最拟合的直

线$y=ax+b$应该与这8组样本数据点距离都很近，最好的情况是这些样本数据点都在该直线上。让直线穿过所有样本点显然不现实，但是可以让所有样本数据点离直线尽可能的“近”。这里的“近”被定义为预测数值和实际数值之间的差。

一旦给定了参数a和b，通过计算$ax+b$得到的值记为$\tilde{y}=ax+b$，接着计算y和$\tilde{y}$之间差的绝对值$|y-\tilde{y}|$，将这个差的绝对值作为x所对应的真实值（即y）和模型预测值（即$\tilde{y}$）之间的误差，这个误差通常称为“残差”。

为了计算方便，在实际中一般使用$(y-\tilde{y})^2$而不是$|y-\tilde{y}|$作为“残差”。这样对于给定的n组(x, y)数据，可用不同的a和b来刻画这n组数据所隐含的$y=ax+b$关系。对于这些不同的参数，最佳回归模型是最小化残差平方和的均值，即要求n组(x, y)数据得到的残差平均值$\frac{1}{n}\sum(y-\tilde{y})^2$最小。

从残差的定义可看出，残差平均值最小只与参数a和b有关，最优解即使得残差最小所对应的a和b的值。

给定8组(x, y)数据，可通过最小二乘法（least square）来求解使得残差最小的a和b。

8组(x, y)样本数据点记为(x_1, y_1)，(x_2, y_2)，…，(x_8, y_8)，时间年份变量x的平均值记为$\bar{x}=\frac{x_1+x_2+\cdots+x_8}{8}$，因变量$y$的平均值记为$\bar{y}=\frac{y_1+y_2+\cdots+y_8}{8}$，那么$a$和$b$值的计算公式如下：

$$a=\frac{x_1y_1+x_2y_2+\cdots+x_8y_8-8\bar{x}\bar{y}}{x_1^2+x_2^2+\cdots+x_8^2-8\bar{x}^2}=1.5344$$

$$b=\bar{y}-a\bar{x}=-2698.9$$

这样，预测莫纳罗亚山地区二氧化碳浓度的一元线性回归模型为：

二氧化碳浓度$=1.5344\times$时间年份-2698.9，即$y=1.5344x-2698.9$。

拓展链接

最小二乘法

最小二乘法是一种机器学习的优化技术，其将残差平方之和最小化作为目标，找到最优模型来拟合已知的观测数据，使得模型所预测的数据与实际数据之间误差的平方和最小，一般有线性最小二乘法和非线性最小二乘法两种方法。用线性最小二乘法来解决线性回归模型存在封闭形式（closed-form solution）唯一解，这个解得到的回归模型使得所有观测数据都在一条直线上或直线附近。非线性最小二乘法需要用数值方法来求解，比如随机梯度下降或者牛顿法等。

可见给定一组数据，在假设数据中存在一元线性关系前提下，就可计算得到一元线性模型来“拟合”这些数据，然后进行预测。表2.5.2在所得到的回归模型基础上，对莫纳罗亚山地区1970年之前和2005年之后的二氧化碳浓度进行估算。

表 2.5.2　莫纳罗亚山1960、1965、2010、2015年的二氧化碳浓度估算

年份x	1960	1965	1970—2005	2010	2015
CO_2浓度y（ppm）	320.04	325.68	表2.5.1中记录的实际二氧化碳浓度值	389.9	400.83

●●● 例1

前面探究了二氧化碳浓度和年份之间的关系，为了进一步探究地球气温变化与二氧化碳浓度之间的关系，收集了如表2.5.3所示的数据。问：二氧化碳浓度和温度之间有怎样的一元线性关系呢？

表 2.5.3　莫纳罗亚山从1970年到2005年间每5年的二氧化碳浓度以及全球温度（相对于1961—1990的平均温度的增长量，经过平滑处理）

CO_2浓度x（ppm）	325.68	331.15	338.69	345.90	354.19	360.88	369.48	379.67
温度y（℃）	−0.108	−0.082	0.015	0.080	0.149	0.240	0.370	0.420

首先，使用Python中常用的一个可视化工具包matplotlib将当前二氧化碳浓度和温度数据在二维坐标中呈现，观察二氧化碳浓度和温度之间是否符合线性分布规律。参考代码如下：

```
import matplotlib.pyplot as plt
#引入画图使用的工具包matplotlib.pyplot

x = [325.68, 331.15, 338.69, 345.90, 354.19, 360.88, 369.48, 379.67]
y = [−0.108, −0.082, 0.015, 0.080, 0.149, 0.240, 0.370, 0.420]
fig = plt.figure()
plt.xlabel("X")
plt.ylabel("Y")
plt.scatter(x,y,c='r')
plt.show()
```

通过上述代码，可以画出二氧化碳浓度和温度的散点图，如图2.5.2所示。

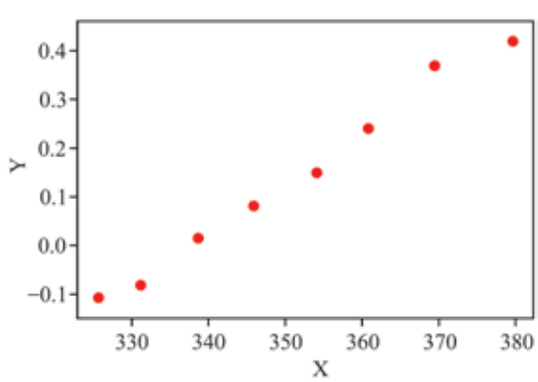


图2.5.2　二氧化碳浓度和温度的散点图

从上面的散点图可以看出二氧化碳浓度和温度之间的分布关系大致是符合$y=ax+b$这样的线性关系。根据最小二乘法计算公式，使用当前给定的数据来计算参数a和b。

在散点图代码之前添加一段计算参数的代码，并在画图步骤中添加画出拟合直线的代码。代码如下：

```
import matplotlib.pyplot as plt
import numpy as np
#引入使用的工具包matplotlib和numpy
x = [325.68, 331.15, 338.69, 345.90, 354.19, 360.88, 369.48, 379.67]
y = [−0.108, −0.082, 0.015, 0.080, 0.149, 0.240, 0.370, 0.420]
x_avarage = sum(x)/len(x)
y_avarage = sum(y)/len(y)                          #计算两个平均数
m = 0
n = 0                                              #两个临时变量用于后续计算参数a和b
for i in range(len(x)):
    m += x[i]*y[i]
    n += x[i]*x[i]
a = ( m−len(x)*x_avarage*y_avarage) / ( n−len(x)*x_avarage*x_avarage)
b = y_avarage – a*x_avarage                        #计算参数a和b
x_predict = np.linspace(325,380,1000)
y_predict = a*x_predict + b                        #按照上面计算出的参数构造直线
fig = plt.figure()
plt.xlabel("X")
plt.ylabel("Y")
plt.scatter(x,y,c='r')
plt.plot(x_predict,y_predict,c='b')
plt.show()
```

通过上述代码，可以画出数据散点图和拟合直线，如图2.5.3所示。

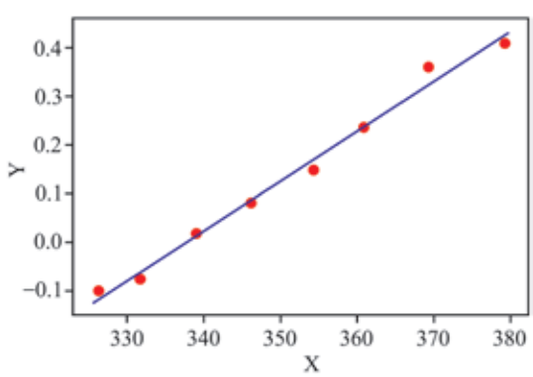


图2.5.3　二氧化碳浓度和温度散点图及拟合直线图

由图2.5.3可以看出，计算所得直线是符合数据点变化趋势的。这说明了最小二乘法在一元线性回归中的有效性。

思考与练习

1. 摄氏温度（°C）和华氏温度（°F）是两种计量温度的标准。表2.5.4给出了两种温度之间的若干关系，如摄氏温度0°C等于华氏温度32°F。

表2.5.4　不同温度下测得摄氏/华氏温度表

摄氏温度（°C）	0	10	15	20	25	30
华氏温度（°F）	32	50	59	68	77	86

试判断摄氏温度和华氏温度之间是否符合线性关系。如符合，请通过回归分析计算出摄氏温度和华氏温度之间的线性回归方程。

2. 摩尔定律（Moore's law）由英特尔创始人之一戈登·摩尔提出，其基本内容为：当价格不变时，集成电路上可容纳元器件的数目，大约每隔18至24个月便会增加一倍，性能也将提升一倍。表2.5.5记录了1971—2004年英特尔微处理器晶体管数量的增长。需要注意的是，随着单位面积上晶体管体积越来越小，摩尔定律所描述的晶体管增长在不久的将来会面临发展极限。

表2.5.5　1971—2004年英特尔微处理器晶体管的数量

微处理器	推出年份x	晶体管数量y	$z=\log_2 y$
4004	1971	2300	11.17
8008	1972	2500	11.29
8080	1974	4500	12.14
8086	1978	29000	14.82
Intel266	1982	134000	17.03
Intel386～processor	1985	275000	18.07
Intel486～processor	1989	1200000	20.19
Intel Pentium processor	1993	3100000	21.56
Intel Pentium II processor	1997	7500000	22.84
Intel Pentium III processor	1999	9500000	23.18
Intel Pentium 4 processor	2000	42000000	25.32
Intel Pentium M processor	2003	77000000	26.198
Intel Pentium D processor	2005	230000000	27.78

注：华为公司2019年研制成功了全球首款名为“麒麟990”的5G芯片。麒麟990集成了103亿个晶体管。

摩尔定律刻画了晶体管数量与时间之间存在指数关系，可用非线性回归拟合来表示这种关系。非线性回归拟合超出了本书的内容范围。不过我们可对晶体管数量取以2为底的对数（记为z），通过判断z与时间x之间是否存在线性关系，来验证摩尔定律（如图2.5.4）。如果上述线性关系存在，使用线性回归方法计算x和z之间的最佳拟合直线。

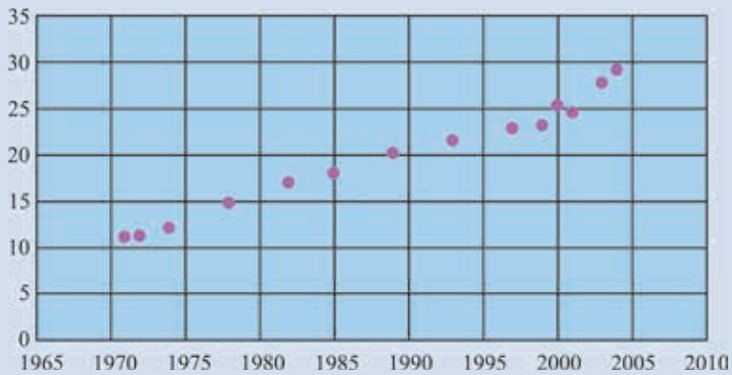


图2.5.4　时间和晶体管数量（取对数后）之间的散点图

2.6 贝叶斯分析

贝叶斯分析是一种根据概率统计知识对数据进行分析的方法，属于统计学分类的范畴，在机器学习中有着举足轻重的地位。本节将介绍贝叶斯分析基础——贝叶斯公式及贝叶斯推断，并通过实例讲解朴素贝叶斯分类器的基本思想。

2.6.1 贝叶斯公式

在现实生活中，常需要估计某个事情发生的概率。例如，A和B是两支球队，在过去十年中，A队和B队一共进行了100场正式比赛，其中A队取得65次胜利。基于这样的事实，可以说A队战胜B队的概率是65%。如果下周A队和B队之间有一场正式比赛，A队战胜B队的概率是多少？基于已发生的历史事实，一般会预测A队下周战胜B队的概率是65%，而不考虑下周出战的A队中有两名主力缺阵、B队新引进了多名高水平外援等新的因素。这一计算概率方法称为频率学派概率，该方法从历史数据中计算某个事件的概率，认为只要采样足够多，则事件发生的频率就可无限逼近真实概率。

与频率学派概率不同，基于贝叶斯框架概率的计算方法认为某个事件发生的概率不仅与先前这个事件发生的概率相关（称为先验概率），而且也与后期计算该事件概率时所观测到的“新近”信息有关（称为似然概率）。

换句话说，在贝叶斯概率计算中，有一个“先验概率”（如A队战胜B队的概率是65%），然后通过观测得到了一些新增信息（如球队主力队员缺阵与否、主场或客场比赛、教练等），分析这些新增信息是增强了“先验概率”还是削弱了“先验概率”，由此得到既考虑了历史事实又重视了当前事实的“后验概率”。

贝叶斯概率计算公式表达如下：

后验概率=先验概率 × 似然概率（即新增信息所带来的调节程度）

拓展链接

“烽火戏诸侯”与贝叶斯思想

《史记·周本记》中有这样的记载：“褒姒不好笑，幽王欲其笑万方，故不笑。幽王为烽燧大鼓，有寇至则举烽火。诸侯悉至，至而无寇，褒姒乃大笑。幽王说之，为数举烽火。其后不信，诸侯益亦不至。”这就是中国历史上著名的“烽火戏诸侯”故事。

在古代，一般“举烽火意味着寇至”（这是先验概率）。周幽王数次举烽火而不见寇至，再次举烽火时，诸侯根据正在发生的事实（如已受骗过几次、周幽王无暇管理国家等）而计算出来的“似然概率”就很小，于是得到了很小的“后验概率”，使得最后真正寇至时举烽火，诸侯计算得到寇至的概率很小了。

“烽火戏诸侯”这个故事说明了事件发生概率与历史有关，也与事件发生的实际情况有关，二者缺一不可。

在介绍贝叶斯定理之前，先解释一下什么是条件概率。条件概率$P(A|B)$表示事件B发生前提下事件A发生的概率。

图2.6.1　A和B之间的交集示意图

由图2.6.1，可直观地计算得到条件概率$P(A|B)$：

$$P(A|B)=\frac{P(A\cap B)}{P(B)}$$

同样地，事件A发生前提下事件B发生的概率可如下计算：

$$P(B|A)=\frac{P(A\cap B)}{P(A)}$$

结合这两个式子，可得到$P(B|A)\cdot P(A)=P(A|B)\cdot P(B)=P(A\cap B)$，于是有：

$$P(A|B)=\frac{P(B|A)P(A)}{P(B)}$$

上面的式子就是有名的贝叶斯公式。其中，$P(A)$是事件A发生的先验概率，与事件B是否发生无关；$P(B|A)$是事件A发生前提下事件B发生的概率，也称为“似然概率”；$P(B)$是事件B发生的先验概率，也称为“标准化常量”；$P(A|B)$是事件B发生前提下事件A发生的概率，也是A的后验概率。

拉普拉斯曾经说，概率论只不过是把常识用数学公式表达了出来。贝叶斯定理就是结合先验常识和似然概率，求取后验概率。

2.6.2　贝叶斯推断

贝叶斯推断是一种基于贝叶斯公式进行分析的统计学方法。为了计算$P(A|B)$，可预先估计一个先验概率$P(A)$，然后根据实际情况来计算$\frac{P(B|A)}{P(B)}$，用其作为调节因子，基于观测得到的新增信息，对先验概率$P(A)$不断修正，得到更为准确的后验概率$P(A|B)$。

下面通过一个广告邮件识别的例子来具体地讲解一下贝叶斯推断。

在互联网产业日益蓬勃发展的今天，很多用户可能都有过邮箱被不感兴趣的广告邮件所充斥的经历。基于贝叶斯推断，现在来设计一个简单的广告邮件自动识别系统。

假设采集到了10000个邮件样本，其中有4000封邮件被认定为广告邮件、6000封被

认定为正常邮件。这4000封广告邮件中，出现“红包”关键词的有1000封，而在6000封正常邮件中仅有6封包含“红包”这个词。

现在，某用户收到了一封如图2.6.2所示的邮件，里面出现了“红包”关键词，那么这封邮件到底是不是广告邮件呢？从已知历史事实来看，它很有可能是广告邮件，如何用贝叶斯推断来计算这个可能性呢？

图2.6.2　广告邮件示意

根据先前收集的10000封邮件，可得到如下计算结果：

用户收到广告邮件的先验概率为$P(\text{广告邮件})=\frac{4000}{10000}=0.4$，

收到正常邮件的先验概率为$P(\text{正常邮件})=\frac{6000}{10000}=0.6$。

广告邮件中出现“红包”关键词的条件概率$P(\text{出现“红包”}|\text{广告邮件})=\frac{1000}{4000}=0.25$，

正常邮件中出现“红包”关键词的条件概率$P(\text{出现“红包”}|\text{正常邮件})=\frac{6}{6000}=0.001$。

根据贝叶斯定理，可得到在邮件中出现“红包”关键词这一前提下，该邮件被认为是正常邮件的后验概率为：

$$P(\text{正常邮件}|\text{出现“红包”})=\frac{P(\text{正常邮件})P(\text{出现“红包”}|\text{正常邮件})}{P(\text{出现“红包”})}=\frac{0.6\times0.001}{1006\div10000}\approx0.006$$

邮件中出现“红包”关键词前提下被判定为广告邮件的后验概率为：

$$P(\text{广告邮件}|\text{出现“红包”})=\frac{P(\text{广告邮件})P(\text{出现“红包”}|\text{广告邮件})}{P(\text{出现“红包”})}=\frac{0.4\times0.25}{1006\div10000}\approx0.994$$

由于$P(\text{广告邮件}|\text{出现“红包”})$远远大于$P(\text{正常邮件}|\text{出现“红包”})$，自然地将该邮件判定为广告邮件。

虽然在人们的“先验”中，邮件中存在“红包”就会被大家以极大概率判定为广告邮件，但是过年期间人们喜欢通过电子邮件拜年，在邮件中会经常出现“红包”单词。

假设在过年期间采集到了10000个邮件样本，其中有4000封邮件被认定为广告邮件、6000封被认定为正常邮件。这4000封广告邮件中，出现“红包”关键词的有1000封，在6000封正常邮件（大多以拜年为主）中有5000封包含了“红包”这个词。

根据贝叶斯定理，在过年期间收到了一封出现“红包”关键词的邮件，该邮件被认为是正常邮件的后验概率为：

$$P(\text{正常邮件}|\text{出现“红包”})=\frac{P(\text{正常邮件})P(\text{出现“红包”}|\text{正常邮件})}{P(\text{出现“红包”})}=\frac{0.6\times\frac{5000}{6000}}{6000\div10000}\approx0.833$$

过年期间邮件中出现“红包”关键词被判定为广告邮件的后验概率为：

$$P(\text{广告邮件}\mid\text{出现“红包”})=\frac{P(\text{广告邮件})P(\text{出现“红包”}\mid\text{广告邮件})}{P(\text{出现“红包”})}=\frac{0.4\times\frac{1000}{4000}}{6000\div10000}\approx0.167$$

可以看出，过年期间邮件中包含了“红包”一词的电子邮件被判定为正常邮件了！

在上述两种情况计算中，先验概率取值是一样的，但是，由于似然概率取值发生了巨大变化，因此产生了很大的调节作用，使得计算结果大相径庭。这进一步说明了判断结果不仅和历史事实有关（先验概率），而且也与当前形势有关（似然概率），两者缺一不可。

另外需要注意的是，在上述后验概率计算过程中，某个事件发生或不发生的后验概率公式中分母均是一样的，因此很多时候在计算后验概率时，我们会省略分母，直接计算“先验概率乘以似然概率”。

在贝叶斯分类中，邮件要被先标注为“正常邮件”和“不正常邮件”，然后开始分类，这种利用了类别标签信息的分类方法，属于有监督机器学习。

2.6.3 朴素贝叶斯分类器

朴素贝叶斯分类器（Naive Bayes Classifier）是一种常用的分类算法，它假设样本各个特征之间相互独立、互不影响。下面通过一个生活中的例子来介绍朴素贝叶斯分类器。

某同学最近下载了一个外卖APP，但看着APP里成百上千家店铺却无从下手，于是在学校论坛上发帖征求外卖美食推荐，接下来的几天他的下单记录如表2.6.1所示。现在，有人向这个同学推荐一家“价位低、口味偏甜、距离远”的店铺，请问他会在该店铺下单吗？

表2.6.1 店铺下单记录

店铺价位	店铺口味	店铺距离	是否下单
高	偏甜	近	是
高	清淡	近	否
高	偏辣	远	否
高	偏甜	远	否
低	偏甜	近	是
低	偏甜	近	是
低	清淡	远	否
低	偏辣	远	是

由表2.6.1可知，该同学在收到8次推荐后，下单4次、没有下单4次。因此，其“下单”和“不下单”的概率计算如下：

$$P(\text{下单})=\frac{4}{8}=0.5$$

$$P(\text{不下单})=\frac{4}{8}=0.5$$

在朴素贝叶斯分类器中，假设店铺价位、口味、距离这些特质之间相互独立、互不影响，因此该同学对这次推荐的“价位低、口味偏甜、距离远”的店铺“下单”或“不下单”的概率为：

$$\begin{aligned}&P(\text{价位=低，口味=偏甜，距离=远}\,|\,\text{下单})\\&=P(\text{价位=低}|\text{下单})\times(\text{口味=偏甜}|\text{下单})\times(\text{距离=远}|\text{下单})\\&=\frac{3}{4}\times\frac{3}{4}\times\frac{1}{4}\approx 0.141\end{aligned}$$

$$\begin{aligned}&P(\text{价位=低，口味=偏甜，距离=远}\,|\,\text{不下单})\\&=P(\text{价位=低}|\text{不下单})\times(\text{口味=偏甜}|\text{不下单})\times(\text{距离=远}|\text{不下单})\\&=\frac{1}{4}\times\frac{1}{4}\times\frac{3}{4}\approx 0.047\end{aligned}$$

根据贝叶斯公式，可以得到该同学在这家价格低、口味偏甜、距离远的店铺下单的概率为：

$$\begin{aligned}&P(\text{下单}\,|\,\text{价位=低，口味=偏甜，距离=远})\\&=P(\text{下单})\times P(\text{价位=低，口味=偏甜，距离=远}|\text{下单})\\&=0.5\times 0.141=0.0705\end{aligned}$$

不下单的概率为：

$$\begin{aligned}&P(\text{不下单}\,|\,\text{价位=低，口味=偏甜，距离=远})\\&=P(\text{不下单})\times P(\text{价位=低，口味=偏甜，距离=远}|\text{不下单})\\&=0.5\times 0.047=0.0235\end{aligned}$$

由结果可见，该同学这次会下单的概率大于不下单的概率。

上面的计算过程进行了一些简化，本来应该计算如下两个公式：

$$\begin{aligned}&P(\text{下单}|\text{价位=低，口味=偏甜，距离=远})\\&=\frac{P(\text{下单})\times P(\text{价位=低，口味=偏甜，距离=远}|\text{下单})}{P(\text{价位=低，口味=偏甜，距离=远})}\end{aligned}$$

$$\begin{aligned}&P(\text{不下单}\,|\,\text{价位=低，口味=偏甜，距离=远})\\&=\frac{P(\text{不下单})\times P(\text{价位=低，口味=偏甜，距离=远}|\text{不下单})}{P(\text{价位=低，口味=偏甜，距离=远})}\end{aligned}$$

上述两个计算公式中分母相同，对计算结果不影响，因此就从计算过程中略去了。朴素贝叶斯假设“价位低、口味偏甜、距离远”是相互独立的，三者同时出现的概率等于三者分别出现的概率，则有如下公式：

$$P(\text{价位=低，口味=偏甜，距离=远})=\frac{4}{8}\times\frac{4}{8}\times\frac{4}{8}=0.125$$

可以想象，如果不是朴素贝叶斯假设“价位低、口味偏甜、距离远”相互独立、互不影响，计算将变得异常复杂。

思考与练习

在表2.6.1中增加一条记录，新表如下所示：

店铺价位	店铺口味	店铺距离	是否下单
高	偏甜	近	是
高	清淡	近	否
高	偏辣	远	否
高	偏甜	远	否
低	偏甜	近	是
低	偏甜	近	是
低	清淡	远	否
低	偏辣	远	是
低	偏甜	远	是

根据新增添的信息，计算“价位高、口味清淡、距离近”这样的推荐被该同学接受并下单的概率。

实践与体验

利用朴素贝叶斯分类器解决MNIST手写体数字识别问题

MNIST是一个手写体数据集，它包含了各种各样的手写体数字图像及其对应的数字标签，可以从官网直接下载IDX格式（一种用来存储向量与多维度矩阵的文件格式）的数据集对其进行解析。

以官网的数据集为例，MNIST数据集由四部分组成，分别为：

（1）Training set images: train-images-idx3-ubyte.gz（训练图像集，包含60000个样本）

（2）Training set labels: train-labels-idx1-ubyte.gz（训练标签集，包含60000个标签）

（3）Test set images: t10k-images-idx3-ubyte.gz（测试图像集，包含10000个样本）

（4）Test set labels: t10k-labels-idx1-ubyte.gz（测试标签集，包含10000个标签）

在MNIST数据集中，每幅手写体图像的大小为28×28，共有784个像素点，可记为一个784维的向量，每个784维向量对应着一个标签。如图2.6.3所示是数据集中的五幅手写体图像，对应的标签分别为6，0，1，1，5。

图2.6.3　MNIST手写体图片示例

MNIST数据集中的训练集和测试集是相互独立的两部分，0—9对应了多种写法，如图2.6.4和图2.6.5所示。

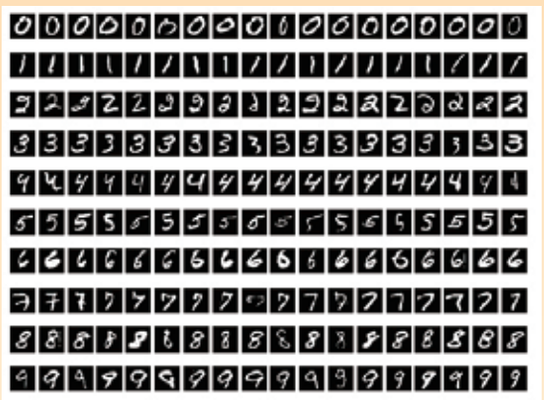


图2.6.4　MNIST训练集图像示例

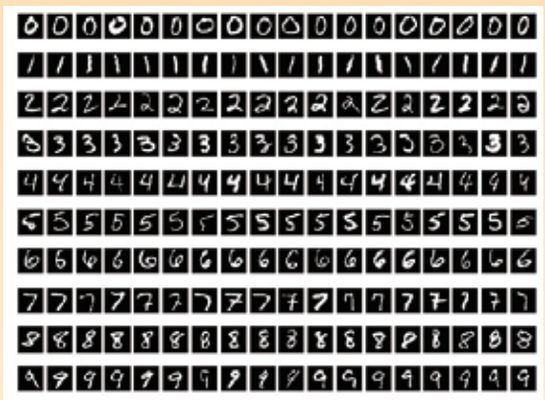


图2.6.5　MNIST测试集图像示例

可以利用Python语言和朴素贝叶斯分类器，从MNIST训练集图像及其对应的标签中学习区别手写体数字的规则，并将学到的规则应用到MNIST测试集图像中，来判断测试集中的每幅图像对应的是0—9中的某个数字。

实践内容：

1. 学会使用Python中的numpy，sklearn和TensorFlow库。
2. 利用朴素贝叶斯分类器解决MNIST手写体数字识别问题。

实践步骤：

1. 安装相关的Python库（以Windows10，Python3.5为例）。

（1）安装numpy库：在cmd控制台输入命令行pip install -U numpy。

（2）安装sklearn库：首先确认计算机中有安装Python（≥2.6或≥3.3版本）、numpy（≥1.6.1版本）、scipy（≥0.9版本），然后在命令行输入pip install -U scikit-learn。

（3）安装TensorFlow库：首先确认Python版本为3.5.x或者3.6.x，然后输入命令行pip install -i https://pypi.tuna.tsinghua.edu.cn/simple --upgrade tensorflow。

在安装结束后可尝试运行import numpy as np、import sklearn、import tensorflow as tf这三条命令，如若不报错则为安装成功。本次实验我们利用TensorFlow库来进行原始数据集的解析和读取，利用sklearn库来进行特征提取和分类。

2. 在Python中导入相应库。

```
import numpy as np
from tensorflow.examples.tutorials.mnist import input_data
from sklearn.naive_bayes import BernoulliNB
```

3. 读取MNIST训练集和测试集。

```
print('reading training data...')
mnist = input_data.read_data_sets("MNIST_data/", one_hot=False)
train_images = np.array(mnist.train.images)
train_labels = np.array(mnist.train.labels)
print('done!')
print('reading test data...')
test_images = np.array(mnist.test.images)
test_labels = np.array(mnist.test.labels)
print('done!')
```

4. 根据MNIST训练集训练朴素贝叶斯分类器。

```
print('Bernoulli Naive Bayes training...')
classifier_BNB = BernoulliNB()
classifier_BNB.fit(train_images, train_labels)
print('done!')
```

5. 根据训练出的分类器对MNIST测试集中的图片进行识别，得到预测值。

```
print('Bernoulli Naive Bayes testing...')
test_predict_BNB = classifier_BNB.predict(test_images)
print('done')
```

6. 将测试图片的预测值与实际值相比较，计算并输出分类器的正确率。

```
accuracy = sum(test_predict_BNB == test_labels) / len(test_labels)
print('The accuracy of Bernoulli Naive Bayes is:', accuracy)
```

结果呈现：

对实验结果进行比较分析，列出0—9不同数字识别的准确率，比较其差异。

2.7 神经网络学习

大脑进行学习的功能基础在于无数神经元之间的连接。每当大脑从外界接触到新的刺激之后，这些神经元之间的连接和状态都发生了改变。这些改变包括神经元之间新的连接产生、强化神经元之间已有连接、去除神经元之间长久未用的连接。神经网络被用来模拟这种学习功能，通过逐层抽象将输入数据映射为概念等高层语义。

2.7.1 人脑神经机制

1981年大卫·修伯尔和托斯坦·维厄瑟尔两位科学家由于发现了“视觉系统分级处理信息”这一机理而获得了诺贝尔生理学或医学奖。比如，当人眼在辨识一幅分辨率为100万像素的汽车图像时，会先从图像中提取汽车边缘特征，然后生成汽车的部件（如轮子、车门等），最后得到更高层的模式。也就是说，高层的特征是低层特征的组合，从低层到高层的特征表示越来越抽象，越来越能表现语义。

人工智能中神经网络（Neural Networks）正是一种体现“逐层抽象、渐进学习”机制的学习模型。今天，机器学习具有从数据中学习的能力，并且用这种能力完成了许许多多任务，如图像识别分类和自然语言翻译，甚至进行绘画、音乐和诗歌等创作。机器究竟是如何一点点来模仿实现的，它们又是通过什么机制来完成的呢?

赫布理论（Hebbian theory）认为神经元之间持续重复的刺激可加强神经元之间的连接，从而强化大脑对这种刺激的“记忆”。

大脑中这些奇妙结构启发了人工智能领域的研究者，他们通过算法来模拟大脑逐层抽象的学习机制。下面通过一个例子来做解释。

图2.7.1 手写体数字示意图

如图2.7.1所示的是手写的一组阿拉伯数字，大多数人毫不费力就可认出这是504192。实际上，大脑在接受了上述信息后，开始调动数以万计的神经元，在这些神经元之间逐层抽象传递信息，完成了上述手写体的识别。

2.7.2 感知机模型

最早的神经网络模型叫感知机（perceptrons）。感知机完成某个任务所使用的思想为“以量取胜”，也就是先收集大量手写体数据（如图2.7.2）作为训练样本（training example），

然后设计一个系统从这些样本中学习出某些模式，最后利用这些模式来识别手写体数字。

那么被称为“感知机”的系统应该是什么？图2.7.3给出了一个感知机模型。

图2.7.2　手写体训练样本

在图2.7.3所示的感知机模型中，有x_1，x_2，x_3三个输入项、一个神经元（用圆圈表示）和一个输出项，每个输入项均通过一定的权重与神经元相连（如w_1是x_1与神经元相连的权重）。当然，在实际中，输入项可以有很多。一旦给定了输入项和权重，神经元如何进行操作得到输出项呢？

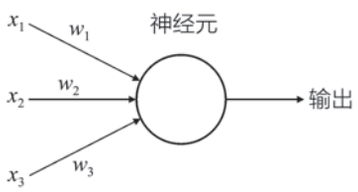


图2.7.3　感知机模型示意图

神经元可采用如下的简单方法来进行工作：先计算输入项传递给神经元的信息总和，即$w_1 \times x_1+w_2 \times x_2+w_3 \times x_3$，如果这个总和大于某个预先设定阈值（比如0.5），那么神经元就输出1，否则就输出0。于是，我们可以写出神经元进行信息处理的数学模型：$F((w_1 \times x_1+w_2 \times x_2+w_3 \times x_3)>0.5)\rightarrow 1$。

也就是说，神经元有两个操作：一是“汇总”与之相连的输入项传递而来的所有信息；二是对汇总后的信息做一个处理（图2.7.3中神经元的处理规则是判断其汇总信息是否大于0.5，然后基于这个判断输出1或0）。

如果有更多的输入，感知机的处理方法也是一样的，只不过需要更多权值w。这就是感知机的工作方法。

思考与练习

1. 如果感知机模型中有n个输入x_1，x_2，…，x_n，请使用数学符号将感知机模型的结果（output）表示出来（假设阈值为0.8）。

2. 请尝试用感知机模型来解决一个现实问题：今天适合出门去郊游吗？（三个决定出门郊游的要素可以为：今天的天气好吗？有同学一起去吗？郊游的地方远吗？）

2.7.3　神经网络

与感知机中输入层和输出层通过神经元直接相连不同，一般神经网络中输入层和输出层之间存在若干隐藏层，每个隐藏层中包含了若干神经元。图2.7.4给出了一个神经网络架构。

在图2.7.4中，有1个输入层，2个隐藏层（hidden layer）和1个输出层。从输入层开始，

第一个隐藏层包含了3个神经元（用圆圈表示）、第二个隐藏层包含了4个神经元。给定某个隐藏层，该隐藏层中的神经元之间没有任何连接。每个神经元与其前后相邻层中的每个神经元均相连。

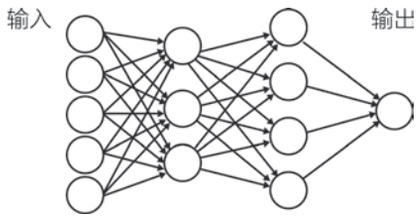


图2.7.4　神经网络架构示意图

在图2.7.4中，第二个隐藏层中的每个神经元接受与之相连第一个隐藏层中所有神经元通过不同权重传递过来的“信息总和”，然后对汇总所得的信息进行处理，再将处理后信息向后传递。

图2.7.4也称为全连接神经网络，即每个神经元均与其前后相邻层中的神经元相连。

可见，一旦给出了图2.7.4神经网络，输入数据从输入层开始，经过两个隐藏层，逐层抽象，最后在输出层得到输出结果。

需要注意的是，图2.7.4神经网络具有如下参数：输入层与第一个隐藏层所有神经元之间的连接权重，第一个隐藏层中神经元和第二个隐藏层中神经元之间的连接权重，第二个隐藏层中神经元与输出层之间的连接权重。

针对某个特定任务（如人脸识别），需要收集与该任务相关的数据来训练神经网络，得到这个神经网络的最佳参数组合，使该神经网络能完成该特定任务。因此，我们可将神经元之间的连接视为“记忆”，即神经网络“记存”了该特定任务的模式。

2.7.4　搭建神经网络

下面我们一起动手搭建一个神经网络，用以解决手写体数字识别问题。

为了方便起见，我们使用现有的手写体数字库MNIST数据集，这个数据集包含了上万张手写体扫描图像以及对这些手写体图像的标注信息。这里推荐Keras这个基于Python语言的深度学习库来完成任务。

首先，为了更加便捷地编程，用import函数导入一些应用函数到程序中。

代码示例

```
import keras
from keras.datasets import mnist
from keras.models import Sequential
from keras.layers.core import Dense, Activation, Dropout
from keras.utils import np_utils
```

接下来，下载所需的MNIST数据集，并将它们转换为模型所能使用的格式。

代码示例

```
(X_train, y_train), (X_test, y_test) = mnist.load_data();
X_train = X_train.reshape(len(X_train), −1)
X_test = X_test.reshape(len(X_test), −1)
X_train = X_train.astype('float32')
```

```
X_test = X_test.astype('float32')
X_train = (X_train−127) / 127
X_test = (X_test−127) / 127
nb_classes = 10
y_train = np_utils.to_categorical(y_train, nb_classes)
y_test = np_utils.to_categorical(y_test, nb_classes)
```

有了这些数据后，需要用到刚刚所学的知识来搭建一个神经网络以解决手写体识别问题。实际上Keras有非常简单的Sequential函数，它能够只用几行代码就搭建两个隐藏层（每个隐藏层中有512个神经元）。

同学们可参考下面的代码来实现手写体识别，其中的Dense函数实现了全连接层功能。这里有一个新的概念，就是激活函数。在图2.7.3中，神经元使用了一个简单的阈值对其汇总的信息进行处理。激活函数是对每个神经元所汇总信息进行处理的一种方法，但是使用了更加复杂的数学模型而不是基于阈值的简单判断，其中比较常见的激活函数有sigmoid、softmax以及relu等等。从激活函数这个名字可以看出，每个神经元并非把所有汇总的信息都往后传递，而是仅仅将“激活”的若干信息向后传递。

代码示例

```
model = Sequential()
model.add(Dense(512,input_shape=(784,), kernel_initializer='he_normal'))
model.add(Activation('relu'))
model.add(Dense(512, kernel_initializer='he_normal'))
model.add(Activation('relu'))
model.add(Dense(nb_classes))
model.add(Activation('softmax'))
```

当这些工作都做完以后，再编写训练和测试神经网络的代码。

代码示例

```
model.compile(optimizer='adam', loss='categorical_crossentropy', metrics=['accuracy'])
model.fit(X_train, y_train, epochs=5, batch_size=64, verbose=1, validation_split=0.05)
loss, accuracy = model.evaluate(X_test, y_test)
print('Test loss:', loss)
print('Accuracy:', accuracy)
```

一旦神经网络搭建完成后，就可用它来完成手写体识别任务，这个神经网络将得到97%的手写体图像识别准确率！

上述程序实现了图2.7.5所示的神经网络，它主要由一个输入层、两个隐藏层和一个输出层构成。

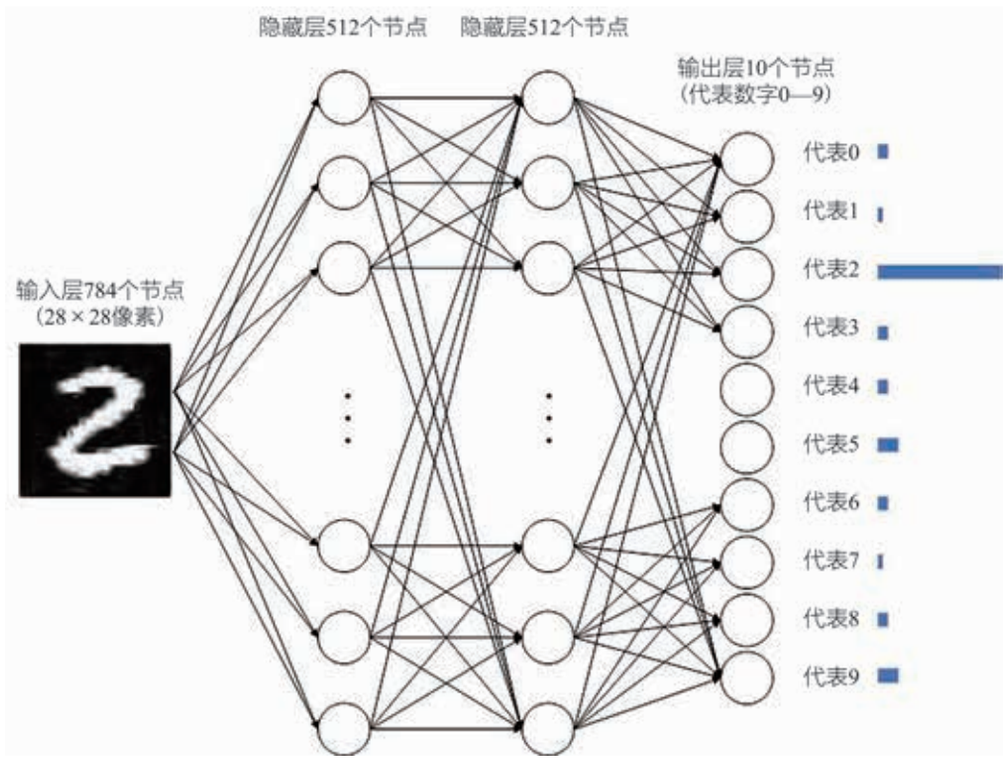


图2.7.5　手写体识别的一个简单神经网络结构图

输入层接受一幅图像的所有像素点信息，输出层是一个10维的向量，每个维度对应0—9这十个数字。对于给定一幅图像，经过图2.7.5神经网络处理后，输出10维向量，我们取10维向量中值最高的那个维度作为识别结果。比如，图2.7.5中输入的图像是“2”，那么所输出10维向量中第三维（注意：第一维对应的是0）概率值应该最大，从而将输入图像识别为2。

拓展链接

赫布（Hebb）可塑性理论（记忆的可能由来）

在巴甫洛夫条件反射中，每次给狗喂食都摇铃，经过多次刺激以后，狗听到铃声就流口水。这一实验说明条件反射具有一定的神经机制。赫布理论所描述的突触可塑性基本原理，在一定程度上揭示了记忆的可能由来。在赫布理论中，突触前神经元向突触后神经元进行持续重复刺激，可导致突触传递效能的增加，即持续重复刺激使得神经元之间“记住”了这种刺激。这是记忆由来的一个猜想。

思考与练习

1. 一般认为，一个神经网络中，隐藏层越多、隐藏层中神经元越多，则这个神经网络的性能越强。请思考这是为什么？

2. 在具体进行神经网络设计时，我们可以设计隐藏层数目多而每个隐藏层中神经元数目少（“瘦高”的神经网络），也可以设计隐藏层数目少而每个隐藏层中神经元数目多（“矮胖”的神经网络），思考这两种神经网络的区别。

实践与体验

调节神经网络结构和参数

在训练神经网络的过程中，选择合适的神经网络结构和参数（如层数）是一个很重要的环节，因为模型所表现出来的效果与这两者具有密不可分的关系。图2.7.5给出了一个包含两层隐藏层的神经网络，每个隐藏层含有512个神经元。

本次实践通过改变神经网络结构来体验不同神经网络结构对手写体数字识别效果的影响。

实践内容：

更改2.7.4小节例子中神经网络结构及其相应参数，观察所训练的神经网络在手写体数字识别中其正确率的变化。

实践步骤：

改变神经网络结构。将2.7.4小节中的两层隐藏层改为一层，模型代码如下所示。模型训练代码与2.7.4小节的代码示例一致，训练模型，在测试集上测试，得出准确率。

```
model = Sequential()
model.add(Dense(512,input_shape=(784,), kernel_initializer='he_normal'))
model.add(Activation('relu'))
model.add(Dense(nb_classes))
model.add(Activation('softmax'))
```

修改2.7.4小节中的两层隐藏层神经元的数量，如将神经元数目设置为2.7.4小节例子中的一半：

```
model = Sequential()
model.add(Dense(256,input_shape=(784,), kernel_initializer='he_normal'))
model.add(Activation('relu'))
model.add(Dense(256, kernel_initializer='he_normal'))
model.add(Activation('relu'))
model.add(Dense(nb_classes))
model.add(Activation('softmax'))
```

训练同上，得出准确率。

结果呈现：

1. 将上述两个实验的准确率与2.7.4小节例子中的准确率进行比较。

2. 输入一个手写体数字，比较三种模型输出结果的差异，对差异进行分析、解释。

2.8 混合增强智能

智能机器与各类智能终端已经成为人类的伴随者，人与智能机器的交互、混合是未来社会的发展形态。人机协同的混合增强智能是新一代人工智能的典型特征。

人类面临的许多问题具有不确定性、脆弱性和开放性，这就需要将人的作用或认知模型引入人工智能系统中，形成混合增强智能形态，这种形态是人工智能或机器智能可行的、重要的成长模式。

出现混合增强智能的原因在于人类智能和机器智能各有优劣，如果能够将这两种智能结合在一起，就可形成既超越人类智能又超越机器智能的新智能体。表2.8.1列出了人类智能和机器智能之间存在的若干区别。

表2.8.1　人类智能和机器智能之间存在的若干差别

人类智能	机器智能
自我学习	从数据样例中学习
自适应	循规蹈矩
具有常识性知识	难以形成
顿悟、直觉、容许不精确	逻辑严密

从表2.8.1可以看出，人类智能和机器智能在自我学习、自适应、常识性知识和顿悟直觉等方面存在差异，这些差异也是机器智能难以完全仿真人类智能的难点所在。

目前，已经出现了一些混合增强智能的例子，如人—车共驾、外骨骼机器人、人机网物相联的城市大脑等。在城市大脑中，个体和群体通过各种APP在城市中留下数字痕迹；物联网感知城市状态（道路流量、天气状况等）；无线网络将这些感知所得的数据传送到云端，云端服务器（机器）对所得到的大数据进行汇聚和处理，进而感知城市脉搏。可见，人、机、网、物等不同智能体各自履行职责，协作完成了城市大脑任务，这是混合增强智能的一种具体体现。

图2.8.1　混合增强智能的例子

图2.8.1给出了不同混合增强智能的例子，如借助壁虎身上所发现的范德华力而制作的爬墙手套、残障人士借助外骨骼机器人踢球、受人类医生指挥的达芬奇手术机器人。

在达芬奇手术机器人中，人类医生凭借高超经验来指挥调动具有工业级精度的手术刀进行复杂外科手术，如果离开了人类医生，机器人无法独立完成这些复杂手术；而没有机器人的帮助，人类医生也无法完成。

思考与练习

查找一种最新的有关混合增强智能的应用，并与同学分享其功能和大致的实现思路，分析在这种混合增强智能中人类智能和生物智能如何协作，从而超越了原有单一智能。

巩固与提高

1. 请写出复合命题$\neg p \wedge (q \vee r)$的真值表。

2. 2.4.2小节中计算了“游客是否前往游乐场”问题中天气状况三个属性值的信息熵以及天气状况的信息增益。请仿照该小节中的计算过程，根据表2.4.1计算温度高低这一气象特点的熵值和信息增益。在计算过程中，将温度高低按照如下取值分为两类：温度大于等于24℃为第一类，温度小于24℃为第二类。根据计算结果，思考“温度”为什么没有出现在图2.4.1所给出的决策树中。

3. 给出如下五组数据，每组数据包含x和y两个值，如表2.8.2所示：

表2.8.2　五组数据

x	−1	1	2	3	4
y	0	2	2	4	5

根据2.5.2小节中给出的公式，可以通过最小二乘法求得拟合这五组数据的直线L_0：$y=0.986x+0.824$。

于是，可求出L_0所对应的残差平方和的均值（以下计算均保留三位有效数字）。

首先，计算L_0对每个x值的预测值：

$y_1=0.986\times(-1)+0.824=-0.162$

$y_2=0.986\times 1+0.824=1.81$

$y_3=0.986\times 2+0.824=2.796$

$y_4=0.986\times 3+0.824=3.782$

$y_5=0.986\times 4+0.824=4.768$

然后，计算残差平方和的均值：

$$\frac{(-0.162-0)^2+(1.81-2)^2+(2.796-2)^2+(3.782-4)^2+(4.768-5)^2}{5}\approx 0.159$$

假如通过另外两条直线L_1：$y=1.1x+0.7$和L_2：$y=0.9x+0.9$来分别拟合这五组数据。根据如图2.8.2所绘制的样本点分布情况和三条直线，能否直观地看出哪条直线“更好地”拟合了这五组数据？试计算L_1和L_2各自对应的残差平方和的均值，并与L_0比较，验证L_0是否更好地拟合了这五组数据。

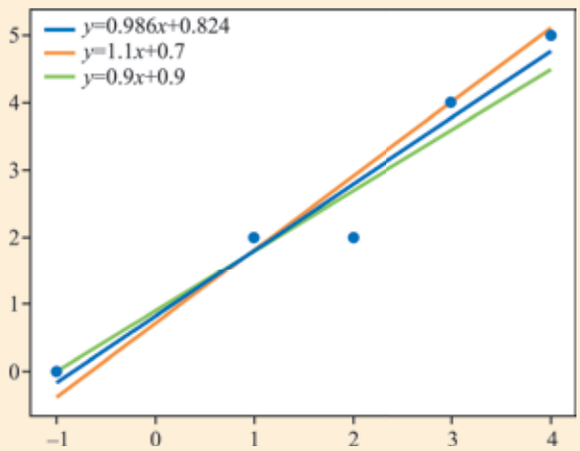


图2.8.2　样本点分布与三条直线的示意图

4. 2.7.2小节介绍了一种感知机模型，神经元对输入项信息以不同权重进行累加，然后根据累加结果是否大于0.5来决定神经元的输出是1还是−1，从而实现了二分类问题求解。在实际中，该模型所实现的二分类问题可用来鉴别人脸与非人脸、疾病与非疾病等。该感知机的模型示意图如图2.8.3所示。

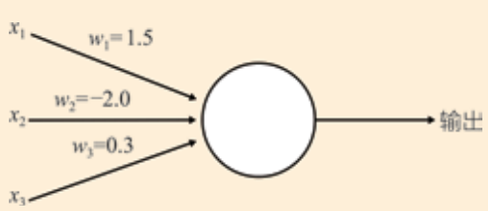


图2.8.3　感知模型示意图

假设有三个样本，x_1，x_2，x_3的值如表2.8.3所示，试计算这三个样本对应的输出分别是多少。

表2.8.3　三个样本

	输入1（x_1）	输入2（x_2）	输入3（x_3）
样本1	1	1	1
样本2	2	1	1
样本3	1	1	0

注：在实际中，一个二分类问题中的输入项不止只有x_1，x_2，x_3三个值。如一个二维的灰度图像数据，可将其输入转换为一个向量数据。

项目挑战

人工智能创新马拉松：物流之链感知城市脉搏

人类社会、信息空间和物理空间正通过前所未有的广度和深度的交互行为深度耦合，形成“人类社会—信息空间—物理空间”三元空间。通过从三元空间中获取的海量数据可洞悉城市脉搏，如从春运大数据来刻画中国劳动力迁移、从手机通话记录来探知城市移民状态、从手机上网活跃程度来计算小区住房空置率等等。这些研究被称为“计算社会学”（computational social sciences），强调通过数据来挖掘可见或不可见的社会生活模式或规律。

中国邮政局发布数据显示，2018年上半年全国累计完成快递业务量达220.8亿件，超过我国2015年全年快递，业务量相当于日本2016年快递业务量，每分钟派送快递量为8.47万件。一件件快递包裹从其发送地到用户接收所形成的“物流链条”蕴含着人类社会中人们工作、生活的信息，对其进行观测分析具有重要意义。

项目任务

以小组合作的形式，收集与物流行业相关的数据。利用本章所学到的知识，对收集到的数据进行分析，试着找出其中的一些规律，并尝试对找到的规律加以解释。

过程与建议

1. 了解计算社会学的背景

通过网络搜索或查询相关资料，了解计算社会学的基础知识。

2. 收集与物流行业相关的数据

在物流公司的网站或其他相关网站收集诸如物流网点的分布、快递包裹收发数量的空间分布等数据；从政府统计网站或其他相关网站收集各省市地区诸如人口数量、经济活跃水平、交通成本等数据。将收集到的数据进行整理，以便后续实验使用。

3. 对收集到的数据进行分析

利用Python sklearn中的工具或其他类似工具，选择合适的方法对收集到的数据加以分析，找出其中的规律。

4. 对得到的规律进行解释

对得到的分析结果展开讨论，尝试解释为什么会有这些规律。对于少数不符合规律的数据，尝试解释其特殊性。

5. 撰写项目报告并展示交流

基于以上工作，撰写分析报告，包括如下内容：

（1）计算社会学的背景。

（2）说明收集到的数据类型，解释希望找到哪些因素之间的关系。

（3）分析数据时使用的方法和得到的结论，以及对结论的讨论。

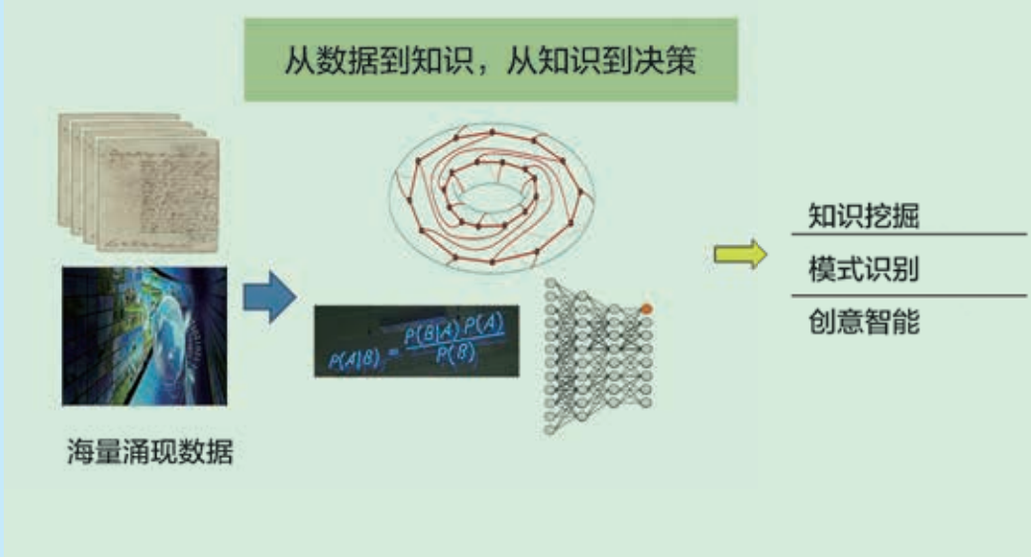
评价标准

请根据项目实施的过程、效果以及成果展示交流的结果，对自己完成项目的情况进行客观的评价，并思考后续完善的方向。把评价结果和完善方案填写在下面的表格中。

评价条目	说明	评分（1～10分）	评分主要依据阐述	后续完善方向
项目理解	在项目开展和分析报告中，展现出对项目目标、内容与任务的正确理解			
小组协作	小组分工合理，协作紧密，合作有成效			
物流相关数据的收集	收集到的数据丰富、可靠			
对数据进行分析	能够获得较为精确的规律			
对规律进行解释	能够解释从数据中得到的规律			
展示交流	展示经过精心准备，表达清晰、便于受众理解			

第三章 3

智能之力：赋能之术



人工智能在人类社会生活中发挥巨大作用的前提是需要与明确的问题场景结合，然后将智能算法赋能于社会生活，从而让人们的生活更加便利。例如，智能算法对顾客消费数据进行挖掘，从而发现对某些商品存在兴趣偏好的用户群；对海量图像进行分析与识别，从而理解图像中的丰富语义；对海量数据进行学习，从而合成新的数据（如自动合成视觉场景）等。

本章主要介绍人工智能赋予人类社会能力的三种代表性技术：知识挖掘、模式识别、创意智能。知识挖掘一般通过聚类方法按照相似度对数据进行组合，使得相似数据能够形成不同聚类集合，从中挖掘相似数据所共有的知识。模式识别则是从数据中学习某个类别的共有模式，基于共有模式进行识别与分类，如人脸识别、行为辨识等。创意智能则是从数据出发，学习数据产生的规律，然后基于这个规律来合成新的数据，如图像合成、乐曲编辑和诗词创作等。

问题与挑战

- 从海量数据中挖掘和发现知识，进而借助知识来指导人类的行为决策（如对市民通过城市公交出行的数据进行感知和理解，优化公交线路等），是人工智能服务社会的一个重要方面，体现了“从数据到知识，从知识到决策”的作用。可以认为一幅图像是数据，从图像中所识别的“行人”“汽车”以及“行人站立在汽车旁”等概念或语义就是知识。思考数据和知识的区别以及知识的表达方式（如通过文本、逻辑或形状等手段表达）。

- 本章介绍了数据聚类和频繁子集挖掘等知识生成的方法。这些方法可从数据中辨析事物之间存在的关联关系（即将下雨和燕子低飞同时发生、面包和牛奶同时被购买等）。大数据中不仅包含事物之间的关联关系，也蕴含事物之间的因果关系（地球绕太阳公转与地球有四季变化）等。因果关系挖掘是人工智能目前的难点。举出一些存在关联关系和因果关系的知识范例，以进一步理解关联关系和因果关系之间的不同。

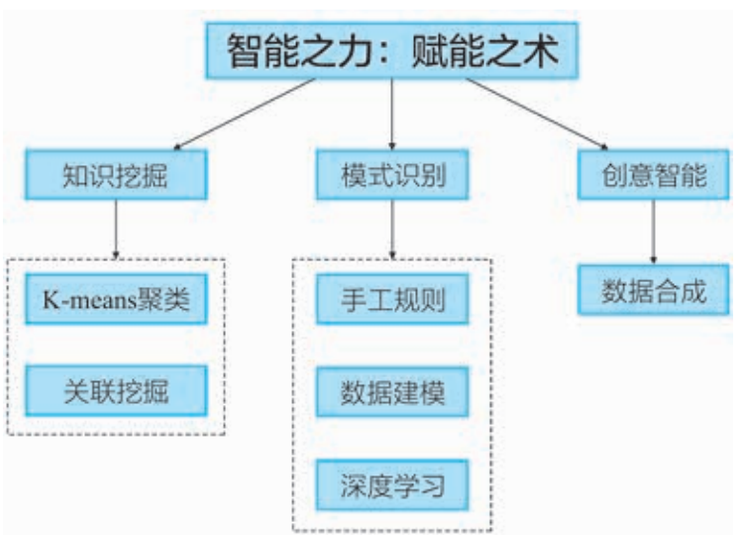
- 深度学习是大数据时代一种有效的人工智能学习方法，这一方法从标注大数据出发，通过逐层递进、不断抽象的方式来获得原始数据表征。深度学习在图像理解、语音识别和自然语言分析等方面取得了显著成效。但是，这一方法依赖于人工标注的海量数据，不可避免具有“人工+智能”的不足，而非“人工智能”。思考深度学习这一方法所适用的任务场景以及不擅长的任务场景。

- 人工智能算法可创造绘画、音乐、海报和诗句等，初步具备了创意能力。人类在创作过程中会融入情感，受阅历等影响，“以诗明志，以画言情”。了解人工智能算法的创意过程，思考这一创意过程与人脑创意的不同。

学习目标

1. 了解数据聚类和关联挖掘等算法模型和应用。
2. 掌握以手工规则、数据建模和深度学习等为核心的模式识别算法和应用及其局限性。
3. 了解人工智能进行数据合成的基本方法。

内容总览



3.1 对数据进行挖掘：知识挖掘

数据挖掘（Data Mining，简称DM）是从海量数据中发现隐性模式或隐含知识的计算过程，它主要用来进行知识发现（Knowledge Discovery in Databases，简称KDD）。

3.1.1 数据挖掘

数据挖掘提供了从数据到知识的一种手段。图3.1.1是数据预处理、数据挖掘和知识应用相互联系的示意图。

图3.1.1　数据预处理、数据挖掘与知识应用过程

数据预处理：数据挖掘是从海量、不完全、含噪声、模糊不确定、随机产生的实际数据中，提取隐含在其中的、人们事先不知道的但是潜在有用的知识的过程。这一过程对数据质量具有一定的要求。因此，需要通过去噪、补全和规整化等手段对数据进行预处理。

数据挖掘：目前有很多从数据中挖掘知识的方法，如聚类方法将相似数据聚集在一起，关联分析从序列数据中发现具有紧密联系的子序列，回归分析建立不同变量之间的联系，分类分析将数据归属到不同类别。其中，聚类和关联分析属于无监督学习方法，回归和分类则因为需要使用到类别信息而属于有监督学习方法。

知识应用：一旦挖掘了知识，则可运用知识进行决策和管理等工作。

3.1.2 数据预处理

从信息世界采集到的海量数据往往存在以下问题：包含噪声信息、若干信息缺失或标准不统一。因此，需要对数据进行预处理，以消除噪声、填补缺失和规范标准。

人口老龄化是当今世界很多国家面临的人口问题，一般一个国家65岁以上人口占总人口的比例达到7%以上，或者60岁以上人口占总人口的比例达到10%以上，就可以判定为进入老龄化社会。

假设某同学收集了一些人口年龄数据（来自联合国2015年人口数据），经过整理得出如表3.1.1所示的若干国家人口老龄化的统计表。

表3.1.1　包含噪声与错误信息的若干国家人口老龄化程度的统计表

国家	人口（千）	60岁以上人口比例（%）	65岁以上人口比例（%）	老龄化程度
日本	127974	32.79	26.02	老龄化
意大利	59504	28.59%	22.36	老龄化
德国	81707	27.35	121.12	老龄化
美国	319929	20.56%	14.64%	老龄化
中国	1397028	15.37%	9.68%	进入老龄化
法国	64457		18.94	老龄化
巴西	205962	11.86	17.96	进入老龄化
埃及	93778	7.71	5.06	老龄化
肯尼亚	47236	4.11	−2.58	未进入老龄化

从表3.1.1可知：（1）数据标准不统一。如60岁以上人口比例（%）和65岁以上人口比例（%）两列中，有些数字后面加了“%”，按照图表的内容描述，不应该在数字后面加“%”。（2）数据存在错误。如德国65岁以上人口比例为121.12%（实为21.12%），显然是一个错误。（3）数据存在缺失。如法国60岁以上人口比例缺失（实际为25.04%）。

这些问题将使得数据挖掘算法难以有效地发挥作用。为了消除这些影响，有些错误可通过对数据的约束性检查来发现，如德国65岁以上人口比例超过了100%、巴西65岁以上人口比例为17.96%（超过了60岁以上人口比例，实际为7.96%）、埃及被标注为老龄化国家（实际尚未进入老龄化）、肯尼亚65岁以上人口比例为负数（实际为2.58%）等。但是，更多的噪声数据或错误数据则需要数据输入进行有效处理以保证数据质量。

下面介绍一种数据预处理的方法：数据归一化（normalization）。

数据归一化操作是一种广泛使用的数据预处理操作。例如，对于一个学生信息表，其身高属性的取值范围可能在1至2 m之间，而其体重属性的取值范围可能在30至100 kg之间，身高和体重不在一个数量级内，不便于后续的数据挖掘。在数据挖掘中，通常需要将数据属性值（如表3.1.1中的人口总数）规整到[0，1]的范围。可采取的一种方法为极大极小值归一化（min-max normalization），公式如下：

$$v'=\frac{v-\min}{\max-\min}$$

其中，v表示原始数据值，v'表示归一化后数据值，min和max分别表示原始数据中最小值和最大值。上述公式将数据取值范围规整到[0, 1]区间。如将表3.1.1中意大利人口总数规整到[0，1]区间可如下计算：

$$\frac{59504-47236}{1397028-47236}\approx 0.009$$

3.1.3　K-means聚类

《战国策·齐策三》中提到，“夫鸟同翼者而聚居，兽同足者而俱行”。后来逐渐演化

为成语“物以类聚，人以群分”。聚类是数据挖掘中的重要概念，其主要思想是按照某个特定的标准（如距离）将一个数据集划分为不同的簇（cluster），使得同一个簇的数据的相似性尽可能大，不同簇的数据的差异性尽可能大。

为了将相似数据放入同一个簇，需要定义数据之间的相似度，即度量数据之间的距离（distance）。

给定数据x_i和x_j，其中$x_i=(x_{i1}, x_{i2},\cdots, x_{ip})$和$x_j=(x_{j1}, x_{j2},\cdots, x_{jp})$分别是两个$p$维数据。计算这两个数据之间距离的常用度量公式有：

（1）曼哈顿距离（Manhattan distance）

$$d(x_i, x_j)=|x_{i1}-x_{j1}|+|x_{i2}-x_{j2}|+\cdots+|x_{ip}-x_{jp}|$$

（2）欧氏距离（Euclidean distance）

$$d(x_i, x_j)=\sqrt{|x_{i1}-x_{j1}|^2+|x_{i2}-x_{j2}|^2+\cdots+|x_{ip}-x_{jp}|^2}$$

曼哈顿距离为方格线距离；在二维空间中，欧氏距离为两个点之间的坐标距离。

拓展链接

曼哈顿距离

曼哈顿距离是由19世纪的赫尔曼·闵可夫斯基所创的，又称为方格线距离。它的命名缘由：在规划为方形建筑区块的城市（如曼哈顿）中，寻找最短的行车路径。如图3.1.2所示，在两黑点间的四条线路中，红、黄、蓝三条线路所表示的曼哈顿距离拥有一样的长度。该距离可计算车辆在城市街道中的行驶距离。

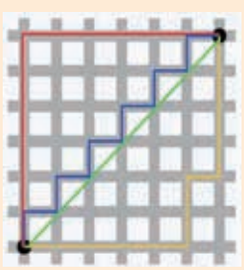


图3.1.2　曼哈顿距离示意图

聚类的方法很多，其中最为经典的是K-means聚类，其算法流程如下：

①从数据集中随机地选择K个数据点作为初始簇中心；

②对剩余的数据，计算其与各个簇中心之间的距离，将其归属到与其距离最近的簇所在的集合；

③给定每个聚类集合，计算其包含数据的平均值，作为该集合新的簇中心；

④不断重复步骤②③，直到集合簇中心不再变化或者变化很小。

K-means聚类具有如下优点：简单，只需要设置一个簇的参数K；高效，在大型数据集上可以并行，有较快的处理速度。但是，K-means聚类也存在如下缺点：需要提前指定K值，初始点的选取会影响聚类结果以及容易忽视属于小类别的数据等。

在商业中，聚类分析常常被用来发现不同的客户群，并且通过购买模式刻画不同客户群的特征，从而进行个性化的服务和推荐。聚类分析也是细分市场的有效工具，可用于研究消费者行为，寻找新的潜在市场。在生物学科中，聚类分析被用于动植物分类和基因分类，从而获取对种群固有结构的认识。在保险行业中，聚类分析通过平均消费来对汽车保

险单持有者进行分组，同时根据住宅类型、价值、地理位置对一个城市的房产进行分组。在互联网应用上，聚类分析被用于文档分类。

思考与练习

给定数据(3,6)和(7,10)，分别计算这两个数据之间的曼哈顿距离和欧式距离。

3.1.4 频繁子集挖掘

在数据挖掘中，发现数据中频繁出现的信息的过程称为频繁子集挖掘。频繁子集挖掘，即挖掘信息之间的关联性，比如商店想要知道哪些货品组合经常被顾客购买，以此来决定商品的进货量、进货时机和商品摆放等。

频繁子集挖掘通常依赖以集合形式存在的数据，比如商店中某顾客一次性购买的所有商品的集合。一旦获得这些数据，频繁子集挖掘就要从这些数据中辨析出出现最为频繁的组合项集。

例如，某商店整理了一天内4名不同顾客购买的商品数据，如表3.1.2所示，希望通过频繁子集挖掘找到商品之间的相关性信息。

表3.1.2 顾客购买的货品信息

顾客	一次性购买商品集合
A	咖啡，牛奶，水果
B	面包，牛奶，果酱
C	咖啡，面包，牛奶，果酱
D	面包，果酱

频繁子集挖掘可以通过很多算法来完成，这里介绍一种典型的算法Apriori。Apriori算法通过不断扩大项集的大小来寻找频繁子集以学习关联规则。

在应用Apriori算法对上述数据进行分析前，先介绍以下几个相关概念。

- K-项集：指包含*K*个项的项集，如*A*购买的商品集合是一个3-项集。
- 项集的频率：出现该项集的次数，又称支持度计数或计数，如{面包，果酱}被三名顾客买过，该项集的频率为3。
- 频繁项集：如果项集的出现频率大于或等于最小支持度计数阈值（由人为设定），那么称它为频繁项集。

这里设定最小支持度计数阈值为2，使用Apriori算法对上述数据进行分析的过程如下：

首先扫描购买信息，找到所有的1-项集，计算它们的出现频率，并与最小支持度计数阈值进行比较，排除小于最小支持度计数阈值的项集，将剩下的1-项集两两之间进行组

合，形成2-项集，计算它们的出现频率，结果如图3.1.3所示。

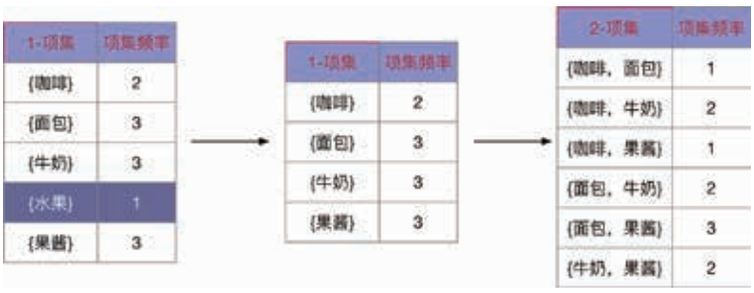


图3.1.3　Apriori算法对商店购物信息分析——1-项集到2-项集

将所有2-项集的频率与最小支持度计数阈值进行比较，排除小于最小支持度计数阈值的项集，将剩下的2-项集两两之间进行组合，形成3-项集，结果如图3.1.4所示。

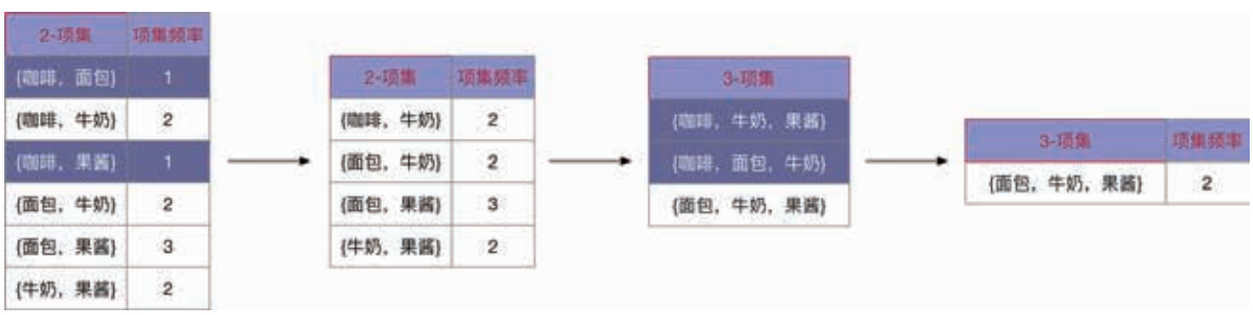


图3.1.4　Apriori算法对商店购物信息分析——2-项集到3-项集

在所有3-项集的集合中，{咖啡，牛奶，果酱}和{咖啡，面包，牛奶}都包含非频繁项集的2-项集子集{咖啡，果酱}和{咖啡，面包}。所以，可以在进一步寻找频繁出现子集前将其删除，接着对3-项集查找项集频率，得到3-项集频繁项集，由于无法组成4-项集，算法终止。

于是，算法挖掘出{面包，牛奶，果酱}是频繁子集。

问题与讨论

如果将表3.1.2中顾客购买的货品信息交换先后位置，如*A*顾客购买的“咖啡，牛奶，水果”变成“水果，牛奶，咖啡”等，这对频繁子集挖掘结果有影响吗？请与同学讨论。

3.1.5　实例分析

给定n个数据样本，每个样本是一个p维的向量，为了更好地可视化，取p=2。图3.1.5给出了7个坐标系的点，采用Python的机器学习工具包scikit-learn对其进行K-means聚类分析，首先需要进行scikit-learn的安装，在命令行中输入：

```
pip install -U scikit-learn
```

安装完scikit-learn以后，就可以调用其K-means函数进行聚类分析了，具体实现如图3.1.6所示。其中，第12行、14行、和16行后半部分的注释表示程序的运行结果。

```
1 #!/usr/bin/python
2 # coding=utf-8
3
4 # 导入必要的工具包scikit-learn和numpy
5 from sklearn.cluster import KMeans
6 import numpy as np
7 # 待聚类的原始数据
8 X = np.array([[1, 1], [1, 2], [2, 2], [2, 3], [4, 2], [4, 4], [5, 3]])
9 # 使用KMeans函数聚成2个类
10 kmeans = KMeans(n_clusters=2).fit(X)
11 # 输出每个数据点的类别
12 print(kmeans.labels_)                    # [0 0 0 0 1 1 1]
13 # 预测两个新数据点的类别
14 print(kmeans.predict([[0, 0], [4, 5]]))  # [0 1]
15 # 输出各个聚类中心的坐标
16 print(kmeans.cluster_centers_)           # [[1.5 2], [4.33333333  3]]
```

图3.1.6　采用scikit-learn工具包进行K-means聚类分析

将图3.1.6的输出结果可视化，可得如图3.1.7所示的结果，其中红色和蓝色分别表示聚出来的两个不同的类，五角星分别表示其聚类中心。

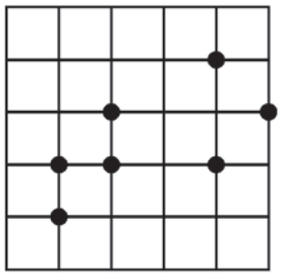


图3.1.5　原始待聚类数据

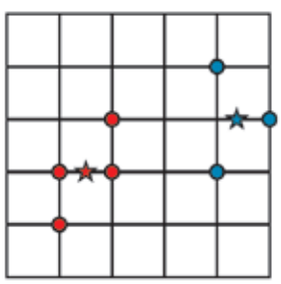


图3.1.7　K-means聚类结果

实践与体验

对数据进行聚类

学习了聚类分析的方法之后，尝试对自己构造或找到的数据进行聚类（为了方便可视化，推荐每个数据点的维度为2），看看聚类结果是否与预期相符，通过实践找出与预期不符的地方。

实践内容：

1. 收集或构造数据。
2. 运行算法。
3. 收集结果，分析与解释。

实践步骤：

1. 收集或构造数据。

收集现实生活中的数据，或对图3.1.6中的注释“待聚类的原始数据”下面代码中的数据进行修改（可以增加、减少或更改数据）。

2. 运行算法。

在运行算法前，你可以对最后聚类结果中的类别数进行修改，修改代码“kmeans = KMeans(n_clusters=2).fit(X)”中的n_clusters后面的数值即可。

结果呈现：

收集实验结果，画出类似图3.1.7的结果图，观察聚类结果是否符合预期，通过实验对不符合预期的结果进行解释。

3.2 对数据进行学习：模式识别

人工智能中的模式识别方法是从数据中学习某个类别的共有模式，然后基于该共有模式实现数据分类。

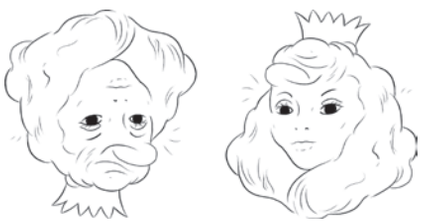


图3.2.1　同一幅图像因颠倒而产生了老妇人和年轻女子两种不同类别

那么，什么是一个数据所包含的“模式”呢？图3.2.1给出了一幅图像及其反转图像。对于这两幅图像（实为一幅），会将左图像及其反转图像分别分类为老妇人和年轻女子两个类别。人们在分类时，首先根据大脑中已形成的“模式”（如眼睛左右对称、鼻子在眼睛下面、闭合嘴巴为弧线等局部结构），然后协调这些局部结构，形成人脸这一全局结构的认知和分类。

显然，人脑在对图像识别分类时，不存在“分类辨识中要先将图像上下颠倒，再甄别和协调局部结构”这样的认知规则，因此一般不会辨识出这两幅图像属于同一幅图像。

人脑可识别出人脸的各个组成部分，那么计算机算法如何做到这一点呢？眼睛、鼻子等人脸组成部分均由若干像素点通过一定的空间布局来构成（如睁开的眼睛可为上下弧线组合），因此，如何从图像中检测像素点不同的空间布局就是要解决的首要问题。

本节主要介绍从数据中学习组成特定对象的模式，进而根据这些模式来对数据进行分类。

3.2.1　基于手工规则的模式检测

早在19世纪就有很多计算机科学家对人脸的重要区域（如眼睛、嘴巴等）的检测进行了研究。那时，科学家依赖的方法是“手工构造”一些规则，进行人脸重要组成部分的模式识别。人脸最为重要的组成部分是眼睛，因此如果能够先准确定位眼睛，那么就能为后续检测出人脸其他组成部分提供线索与帮助。

人眼由黑色像素点组成，构成人眼的像素点比周围像素点的取值要小。于是，在一幅人脸图像中寻找局部最小值（取值比周围像素点值都要小）的像素点。在找到这样的像素点后，这个局部最小值所在像素点区域会以极大概率被识别为眼睛。接着，在眼睛区域附近根据相关规则（如对称等）继续搜索其他人脸组成部分。

图3.2.2给出了人脸组成部分检测的示意图，其展示了先检测眼睛，再检测鼻子等局部结构，最终协调这些局部结构来形成全局结构（人脸）。

但是，这种基于规则方法的检测手段并不可靠。例如，某人脸图像中，如果双眼紧闭或者戴着一副茶色眼镜，那么就很难率先检测到人眼，进而检测其他人脸组成部分。

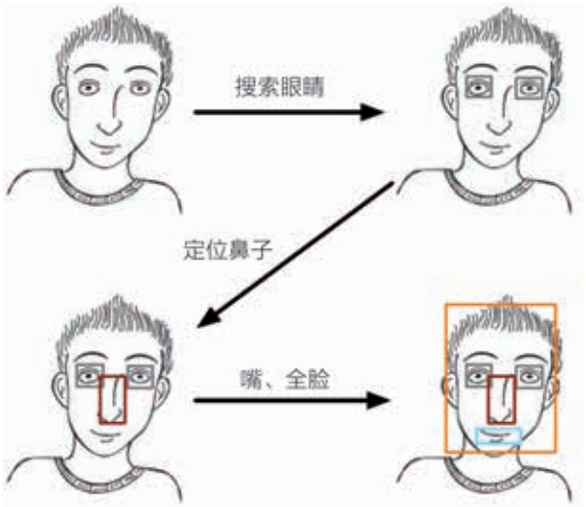


图3.2.2　基于规则的人脸组成部分检测结果

在上述算法中，人脸组成部分的检测首先基于眼睛位置能被正确检测，再检测其他人脸器官。而后在前一轮检测的基础上再开始下一轮检测，这样一连串“前后相互依赖”检测势必导致“一步走错，满盘皆输”的窘况。另外，繁杂冗长的检测规则也难以迁移到其他类似问题的解决中，如人脸检测规则难以直接用于车辆检测。

3.2.2　基于数据建模的模式检测

除了依靠规则进行人脸模式检测外，也可通过数据建模来实现人脸中内嵌模式的检测。

如图3.2.3所示，定义了一个由很多关键点组成的人脸模型。从这个人脸模型的定义中，能看到眼睛、嘴巴、鼻子等组成部件以及它们之间形成的几何约束，如眼睛左右对称、嘴巴在鼻子下方等。当然，与基于手工规则检测方法不同，这些约束并非人为事先定义，而是从数据中直接学习。从数据出发直接学习模式与事先构造识别规则之间有着重要区别。

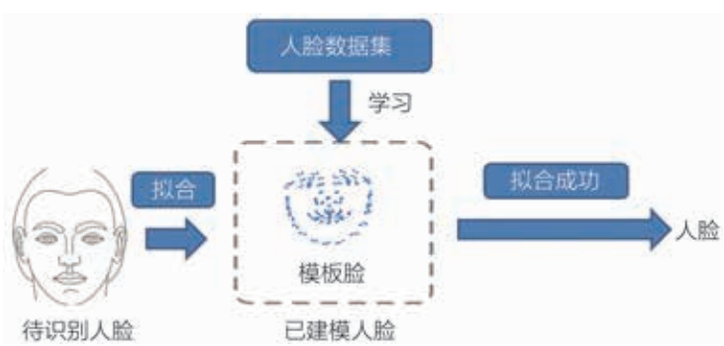


图3.2.3　基于数据建模的人脸识别

在基于数据建模的模式检测中，先要从数据中对一个或多个模式进行建模，然后设计算法从图像像素点中“拟合”出与先前学习得到的人脸模式一致的人脸模型。目前，这类方法的代表性算法是ASM（Active Shape Model）和CLM（Constrained Local Model）等。

下面介绍通过数据建模的方法来进行人脸识别。先给定一幅人脸图像，从该图像中检

测人脸关键点，这些关键点按序拼接起来可表示人脸轮廓（如图3.2.3所示）。然后，给出另外一幅人脸图像，按照同样的方法检测出人脸关键点。如此循环，直到从所收集人脸数据集中均检测得到了人脸关键点。

由于每幅人脸图像中的人脸形态存在差异性，加上拍摄环境不同（光照强弱、距离远近不同等），因此从每幅人脸图像中所提取的人脸关键点也存在差异性。为此，接下来从所有人脸关键点出发，学习拟合一张“模板脸”，这张模板脸可反映人脸的差异性，称这张模板脸为“均值人脸”。

得到均值人脸后，任意一张人脸都可以由均值人脸通过一系列参数变化得到，也就是说在特定参数控制下，均值人脸可以“变形”生成每张独特的人脸。

一旦训练得到了均值人脸，测试一幅图像是否为人脸，只需检测这幅图像中的关键点是否与均值人脸拟合。若拟合，则认为该图像中存在人脸；否则，不存在人脸。

总的来说，基于数据建模的模式检测方法所需的数学知识和实现的复杂程度比基于规则算法更多、更难，当然，检测效果也更加显著。并且，基于数据建模的方法不需要预先定义规则，而是直接学习数据中所蕴含的模式，这使得算法更灵活。

3.2.3 基于深度学习的模式检测

深度学习通过逐层抽象方式来学习数据中蕴含的模式。与数据建模的“浅层计算”方法不同，这种“深度”学习模型具有更强的表达能力，在大数据驱动人工智能背景下取得了显著成功。

深度学习机理可追溯到早期神经科学的发展。1958年，修伯尔和维厄瑟尔两位科学家在进行猫视觉皮层实验时，首次观察到了位于视觉初级皮层的神经元对移动的边缘刺激敏感，发现了视功能柱结构，解码了大脑对眼睛所发出信息的解读秘密。这一工作为视觉神经研究奠定了重要的基础。

在20世纪50年代，感知机这一模型被提出。感知机是一个只有输入层和输出层的神经网络。明斯基在1969年指出感知机无法用来识别复杂的模式。

近年来，计算机计算能力有了大幅的提升，互联网的发展使得海量数据不断涌现。在这样的背景下，借助海量标注大数据，通过学习模型，在超强计算能力的帮助下来训练几十层甚至几百层的神经网络成为可能，推动了深度学习的蓬勃发展。深度学习为计算机视觉、自然语言处理中的许多问题带来了突破，例如在图像识别和分类等任务中取得极大成功的卷积神经网络（Convolutional Neural Network，简称CNN），正是一种受到人类视觉系统信息加工机理启发的深度神经网络。

目前，有一些基于深度学习算法开发的识别分类产品或APP，比如能识别语音内容并进行相应操作的华为智能音箱、能听歌识曲的软件《网易云音乐》，以及能够识别多种植物种类、快速给出植物名称的花草识别软件《形色》等。

1. 卷积滤波

卷积神经网络中最为重要的概念是“卷积”（convolution）。下面通过图3.2.4来解释卷积的基本概念。在图3.2.4中，滤波器（filter）用来对图像进行处理。这个3×3滤波器从上到下、从左到右的9个值为（1, 2, 1, 2, 4, 2, 1, 2, 1），这9个值称为滤波器的权重参数。

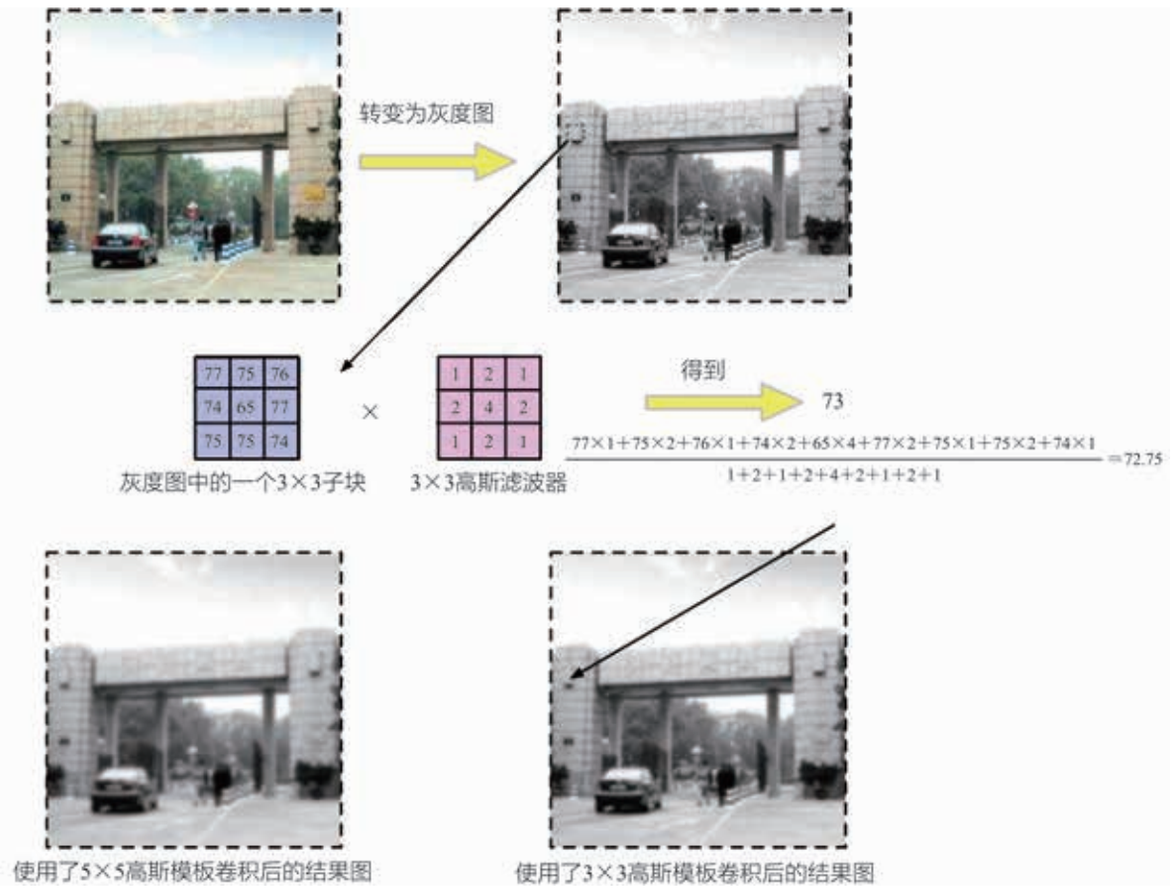


图3.2.4　图像卷积操作示意图

在图像滤波中，将被滤波图像中的某个像素点作为中心像素点，形成一个与滤波器大小一样的3×3分辨率图像子块，然后将3×3滤波器中每个方格位置的值与3×3图像子块相应位置的值相乘，并将乘积累加后除以3×3滤波器所有权重之和，得到的结果作为该中心像素点的值。例如，图3.2.4中“65”这个中心像素点被3×3滤波器滤波后，其值变为73。图像中所有像素点都这样处理后，得到的结果称为图像滤波后结果，即卷积后结果。当然，以图像中的边界像素点为中心无法形成3×3的子块，因此，边界像素点不参与卷积计算。

在卷积计算中，如果卷积矩阵中的值均为1，那么中心像素点卷积后的值等于其与邻域8个像素点的平均值。如果中心像素点与邻域像素点之间存在较大的差别，那么这种差别经过卷积后会减少。也就是说，卷积后图像中像素点取值会趋同，这时称图像被平滑滤波了。如图3.2.4中的中心像素点“65”被滤波后变为73，其值与领域像素点靠近。平滑滤波的效果是图像变得模糊。图3.2.4中给出了通过两个高斯滤波器对原始图像进行平滑

滤波的效果，可见图像变模糊了。

要注意的是，滤波器中的权重参数是通过数据驱动学习得到的，而不是手工指定的。这些训练得到的滤波参数能够将隐藏在图像数据之中的模式提取出来，有利于后续分类。

2. 激活函数

视觉理解的目的是将图像像素点这一信息映射为语义内容，实现从像素点空间到语义空间的一次复杂变换。如图3.2.5所示，在视觉理解中，“非线性映射”将左图图像（像素点集合）映射到高层语义（如冬雪、行人与树木等）。

图3.2.5　视觉理解中的非线性映射：从像素点到语义

深度学习中也应该考虑非线性映射机制，用以增强机器学习的性能。在深度学习中，非线性映射这一机制由激活函数来完成。从神经网络角度来说，激活函数考虑的是什么信息可以被激活而传递给后续神经元、什么信息因其未能激活而无法传递给后续神经元。

在卷积计算中，每个像素点卷积结果为其值与邻接像素点的线性累加。假设像素点的线性累加结果为Con_{pixel}，对累加结果做如下非线性变换：

$$f(Con_{pixel})=\max(Con_{pixel}, 0)$$

这个非线性变换函数叫线性整流函数（Rectified Linear Unit，简称ReLU），其功能是激活线性累加为正的结果。当线性累加结果为负数时，ReLU结果为0（线性累加结果为负数时，该信息不激活）。

这样在对每个像素点进行卷积计算以提取空间分布模式时，同时对卷积计算结果做一次非线性变换，进一步提升学习算法的性能。

使用ReLU作为激活函数，其本身没有需要学习的参数。

3. 池化操作

人脑对外界信息会不断地抽象，去繁就简，如在理解一幅100万像素点构成的人脸图像时，人脑会很快对这100万像素点进行压缩，从中提取出有意义的信息。

深度学习通过池化操作（pooling）将原始信息“由厚变薄”，保留应该保留的信息。如图3.2.6中给出了最大池化（max-pooling）和均值池化（average-pooling）操作的结果。在将原始图像划分为2×2子块后，前者从每个子块中选择一个值最大的像素点，后者从每个子块中选择该子块中像素点均值，来完成原始图像的池化操作。

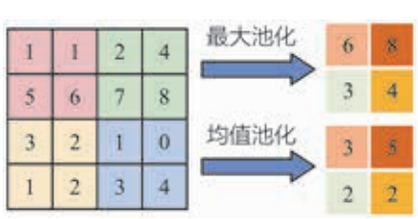


图3.2.6　最大池化和均值池化操作示意图

每个子块中的最大值或均值可以作为该子块代表。因此，池化操作不仅保留了原始图像的有意义信息，也对图像进行了压缩（如图3.2.6所示，图像大小由4×4变成了2×2），

达到了对图像内容进行抽象的效果。

与激活函数ReLU一样，最大池化和均值池化没有需要学习得到的参数。

4. 全连接层

全连接层就是前面讲过的感知机模型，相邻层之间神经元相互连接，将卷积和池化操作所得的结果转变成向量，以便更好地进行分类。全连接层中层与层之间的连接权重均是要训练学习得到。

例1

在学习了卷积层、激活函数、池化层和全连接层后，下面来构建人脸重要组成部分的检测任务中的深度模型。

假设有一个人脸图像数据集，已经在人脸图像中标注了关键点，然后通过标注数据训练了一个深度模型来检测这些关键点。如果给出一张人脸，这个训练好的深度模型会从人脸图像中检测这些关键点。基于这些检测得到的关键点信息，可再进行后续的图像人脸分析。

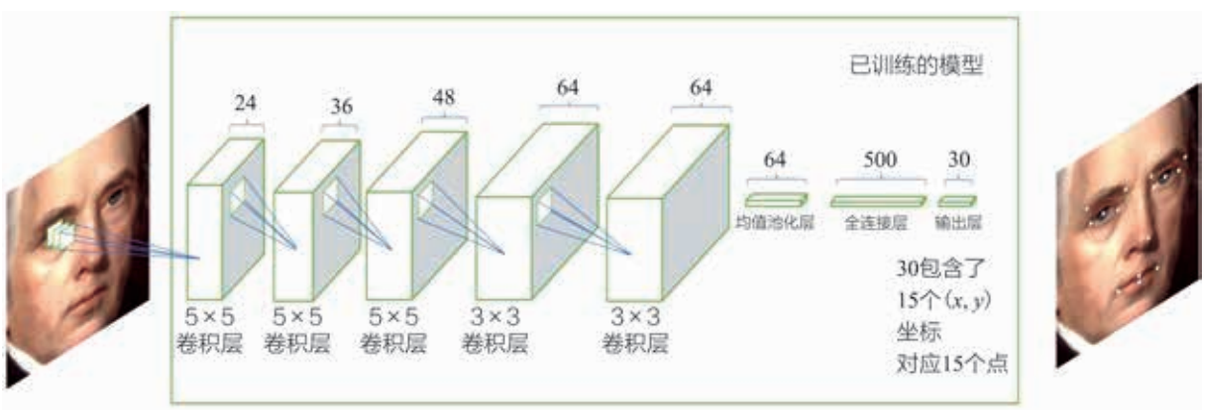


图3.2.7　通过卷积神经网络检测人脸关键点所输出的网络结构

图3.2.7给出了人脸关键点检测的流程图，其包含了五个卷积层。先是三个5×5的卷积层，卷积滤波器个数分别是24，36，48；接着是两个3×3的卷积层，卷积滤波器个数分别是64，64；然后是一个均值池化层；最后是一个全连接层和一个输出层。输出层输出15个关键点的横、纵坐标位置，因此输出层是30维向量。

为什么需要这么多的滤波器？在上面的解释中，我们知道每个滤波器就是按照不同“模式”来挖掘图像像素点与其相邻像素点之间存在的不同的空间分布模式。图像中所包含的视觉内容非常复杂，经过多个滤波器的处理，可以学习得到更多图像像素点的不同分布模式，反映图像像素点的复杂空间变化。

表3.2.1　卷积神经网络学习中不同操作的解释

操作名称	作用	参数学习
卷积计算	挖掘中心像素点与周围像素点之间存在的模式。卷积（滤波器）个数越多，挖掘出的像素点空间模式越多	卷积的权重参数要通过深度学习训练得到
池化操作	保留图像中的重要信息，对图像进行约减抽象	没有需要训练的参数
激活函数	对神经网络进行非线性变换	没有需要训练的参数
全连接层	将特征映射为向量	连接权重主要通过深度学习训练得到

表3.2.1列出了深度学习模型中一种有代表性的模型——卷积神经网络不同操作的作用和参数学习等信息。在标注大数据驱动学习下，一旦训练得到一个良好的深度模型，则该模型可提取数据中最能表达语义内容的特征，然后将这个特征用于分类和识别。因此，深度学习也被称为特征学习（feature learning）。很显然，由于利用了类别标注信息，卷积神经网络这种学习模型是有监督学习。

使用Keras来实现人脸中关键点信息检测的语句也非常简单。“model.add(Convolution2D(64, 3, 3))”这一语句就实现了增加一层卷积层，里面包含了64个3×3卷积滤波器。

下面代码展示了在Keras平台上设计和学习一个神经网络来实现人脸关键点的检测。

```
model = Sequential()
model.add(BatchNormalization(input_shape=(96, 96, 1)))      #输入96×96分辨率大小的图像
model.add(Convolution2D(24, 5, 5, border_mode="same", init='he_normal', input_shape=(96, 96,
1), dim_ordering="tf"))                                     #24个5×5大小的卷积函数
model.add(Activation("relu"))                               #在卷积中采用激活函数
model.add(Convolution2D(36, 5, 5))                          #36个5×5大小的卷积函数
model.add(Activation("relu"))                               #在卷积中采用激活函数
model.add(Convolution2D(48, 5, 5))                          #48个5×5大小的卷积函数
model.add(Activation("relu"))                               #在卷积中采用激活函数
model.add(Convolution2D(64, 3, 3))                          #64个3×3大小的卷积函数
model.add(Activation("relu"))                               #在卷积中采用激活函数
model.add(Convolution2D(64, 3, 3))                          #64个3×3大小的卷积函数
model.add(Activation("relu"))                               #在卷积中采用激活函数
model.add(GlobalAveragePooling2D())                         #池化操作
model.add(Dense(500, activation="relu"))                    #全连接层中采用激活函数
model.add(Dense(30))                                        #输出层
```

上述深度神经网络会输出人脸图像中15个关键点的位置，这15个关键点合起来反映了人脸的不同器官。由于每个像素点有x和y两个坐标值，因此一共输出30个数据。

拓展链接

深度学习模型的训练：误差后向传播

误差后向传播是深度学习中一种训练与学习参数的方法，由人工智能著名学者杰弗里·希尔顿于1986年提出。误差后向传播通过标注大数据驱动的学习方式来优化训练神经网络中的参数，如卷积滤波权重参数和全连接权重参数等。这样，使得深度模型对输入数据的输出结果与该数据的实际标签相一致。可见，深度学习离不开层层抽象深度模型、标注大数据和参数学习算法。因此，有些专家认为“模型、数据和算法”是深度学习的基本元素。

思考与练习

1. 图3.2.7中给出了包含卷积层、池化层和全连接层的卷积神经网络结构。卷积层操作考虑了图像像素点之间的空间相关性，如某个像素点被卷积操作的结果与如下三个因素相关：（1）该像素点的取值；（2）与该像素点相邻像素点的取值；（3）卷积核中的权重大小。在全连接层操作中，像素点的空间相关性被忽略。请从保持像素点空间相关性的这个角度，举一个例子说明，在卷积神经网络中，卷积层操作比全连接层操作更为重要。

2. 更改文中人脸关键点检测的代码，如增加/减少卷积滤波器数目、增加/减少全连接层数目，观察新模型下卷积神经网络的性能变化，并分析原因。

实践与体验

动物图像的分类

本节学习了从数据中学习以及如何对数据进行分类，这次实践将使用Keras对如图3.2.8所示的图像进行分类，让深度学习算法检测图像中的动物是什么。

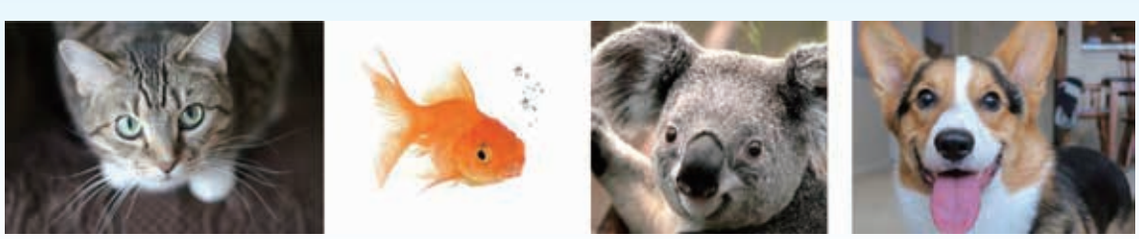


图3.2.8　不同动物的图像

实践内容：

1. 对图像进行预处理。

2. 通过一种深度模型残差网络（Residual Network, 简称ResNet）对图像进行分类。残差网络结构如图3.2.9所示。（注意：此图内容不要求掌握。）

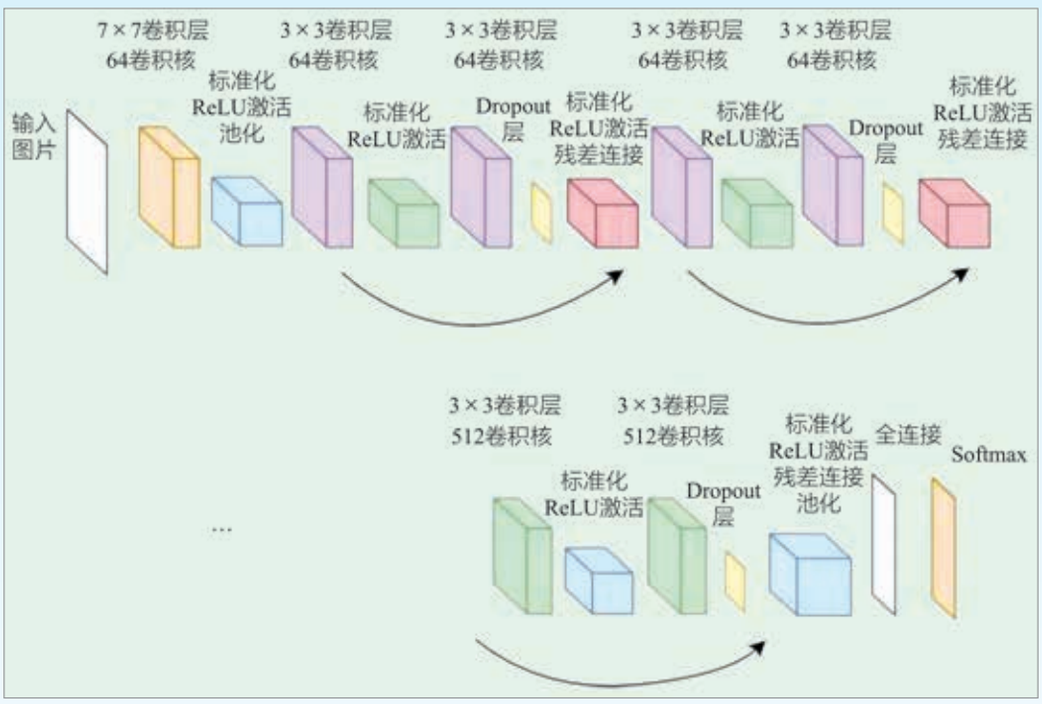


图3.2.9　ResNet模型示意图

实践步骤：

1. 图像预处理。

Keras中ResNet模型的输入图片需要是224×224，手动收集的图片不一定满足条件，需要对图片进行预处理（包括但不限于对过大的图片进行裁剪，对较小的图片补上黑色或白色边框）。这些预处理可以手工处理，也可以利用Python的cv2库进行处理。

cv2的安装和使用操作代码如下：

（1）安装cv2。

在命令行输入：

```
pip install opencv-python
```

（2）使用cv2将图片转换为224×224的代码。

```
import cv2
image = cv2.imread(image_path)                                  #从image_path读取图片
# 将图片转换成224*224
new_image = cv2.resize(image,(224,224))
# 将转换后的图片存放到new_image_path
cv2.imwrite(new_image_path, new_img)
```

2. 通过ResNet模型对图像分类。

```
from keras.applications.resnet50 import ResNet50
from keras.preprocessing import image
from keras.applications.resnet50 import preprocess_input, decode_predictions
import numpy as np

model = ResNet50(weights='imagenet')                            #构建ResNet50 模型

img_path = 'elephant.jpg'
# 将单引号中的内容换为你想要分类的图片路径
img = image.load_img(img_path, target_size=(224, 224))          #加载图片
x = image.img_to_array(img)                                     #将图片转换为数据矩阵
x = np.expand_dims(x, axis=0)
x = preprocess_input(x)

preds = model.predict(x)
#对图片种类进行预测，得到种类分布
#根据种类分布拿到真正种类
print('Predicted:', decode_predictions(preds, top=3)[0])
```

结果呈现：

收集若干张相同或种类相近的动物（如猫、虎等）图像，比较分类结果的不同；收集其他种类（非动物）的图像进行分类，了解本次实验所选模型的局限性。

3.3 对数据进行合成：创意智能

上一节介绍了如何从数据中学习，什么是数据中包含的“模式”，以及如何利用这些“模式”实现对数据的分类和检测等，这种模型通常称为判别模型（Discriminative Model）。

当前，人工智能算法可从数据中学习，生成新的数据，进行诸如诗词创作、画作生成和音乐合成等工作。这一节将介绍生成模型（Generative Model）的基本理念（有别于上一节学习的判别模型），一种有效生成模型——生成对抗网络，以及生成对抗网络模型的若干应用。

3.3.1 生成模型

上一节学习的模式检测和分类都属于判别模型的内容，判别模型不关心数据如何生成，它只关心数据蕴含哪些模式以及如何将数据进行分类。有别于判别模型，生成模型需要掌握如何生成数据。类似地，人在学习过程中，能阅读和理解一篇文章，却不一定能够写出同等水平的文章；能够做海量题目，却难以出一道好题。判别模型和生成模型的区别可以用“授之以鱼”和“授之以渔”来阐述。通常来讲，设计和训练一个生成模型比制作一个判别模型更加困难或更耗费资源。

本质上，生成模型学习数据产生机制的过程就是学习目标数据分布规律的过程，如果掌握了数据分布规律，那么就可以按照所习得的规律从数据空间中抽取样本，从而生成所需的新数据。

在现实应用方面，生成模型可以用于很多领域，比如许多新闻媒体使用生成模型人工智能算法进行体育赛事和股票行情报道、一些壁纸印刷厂商使用自动生成图像纹理来印刷产品、网络商户根据存货内容和销售行情来绘制海报等。

3.3.2 生成对抗网络

生成对抗网络（Generative Adversarial Network，简称GAN）是由伊恩·古德费罗等人于2014年提出的一种生成模型。生成对抗网络由一个生成器（Generator，简称G）和一个判别器（Discriminator，简称D）组成。

GAN核心思想隐含在其名字中——“对抗”，即通过生成器和判别器两个神经网络之间的竞争对抗，不断提升彼此的水平，从而使得生成器最后能够生成判别器难以识别的合成数据。GAN的思想可用名画模仿者和画作鉴定师之间的关系来解读：一开始，名画模仿者的模仿水平和画作鉴定师的鉴定水平都很差。模仿者通过不断学习名画来让鉴定师看不出其创作赝品与真迹的差异，鉴定师也通过对名画和赝品的研究来提升判别名画模仿者

的赝品的能力。在这种竞争对抗中，名画模仿者和画作鉴定师的水平均得到了提升。

下面给出GAN基本算法描述。

生成器（G）所完成功能是将简单的分布（比如正态分布等）转化成符合真实情况的复杂分布，复杂分布中每一个点就是生成目标的一个样本、一幅画、一首诗或一段音乐等。如在手写体数字图像生成问题中，生成网络需要完成的任务就是将从正态分布中抽样所得的采样点转化成为一幅真正手写体数字图像（可以想象这种转化是非常复杂而困难的）。为了达到这一目的，通常选择神经网络来作为生成器（G）的基础架构。神经网络的强大拟合能力使这种复杂转化变为可能。

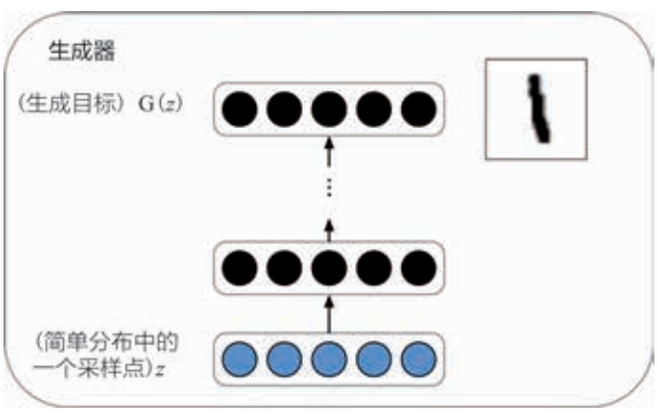


图3.3.1　一种简单的生成器结构图

图3.3.1中的一个样本所示的是一种简单的生成器结构。生成器从简单分布中采样一次，生成复杂分布，本例即合成产生手写体数字1。

判别器（D）的任务是判断所输入图像是由生成器合成产生（用数值0表示，意味着其是合成的虚假数据），还是来自真实数据集（用数值1表示，意味着其是真实数据）。

在训练中，生成器（G）的目标是“骗过”判别器，因此需要得到来自判别器反馈信息以知道“骗过”任务是否完成，即要知道判别器对其生成数据的“评价”（该评价反馈是一个介于0和1之间的数值，越接近于0，表示判别器越倾向于认为这个数据是由生成器合成的）。这些“评价”能帮助生成器不断提高所生成图像的效果，同时，判别器也会在不断识别出更加具有迷惑性的合成图像的过程中提升判别能力。一种简单的判别器如图3.3.2所示。

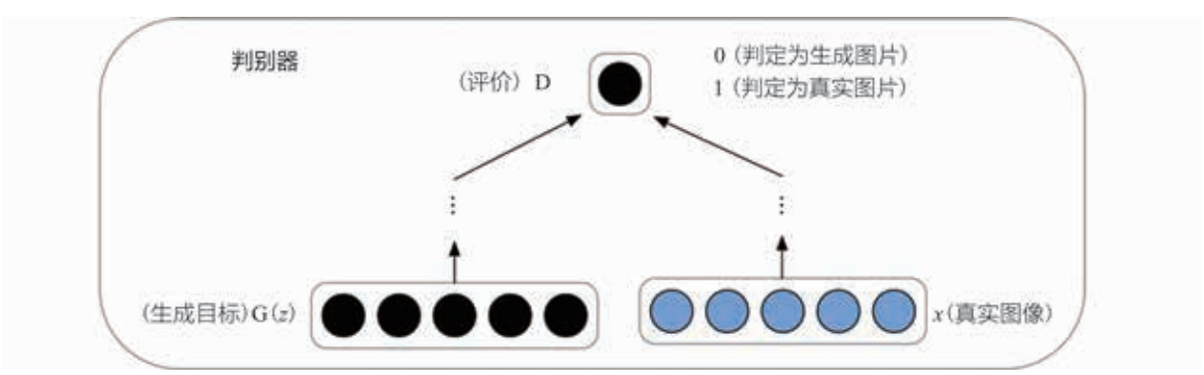


图3.3.2　一种简单的判别器结构图

3.3.3 生成对抗网络及应用

生成对抗网络可应用于许多任务，如多种类的图像生成、文字到图像的转换、图像风格迁移、超分辨率（将低分辨率图像转换成高分辨率图像）以及图像编辑（变换物体某个部分的颜色）等。

图3.3.3所示为中文字体生成和图像风格转换（将真实照片转换成莫奈和梵高的画像）。

图3.3.3　条件生成对抗网络及应用

针对不同的任务，许多生成对抗网络的变体被提出。其中，条件生成对抗网络是一个很重要的研究方向，它被提出用来解决在一定条件下可生成某些任务的问题。简单来讲，条件生成任务就是带有约束的生成任务。条件约束可以是手写数字生成中0—9的数字信息，多种类图像生成中花、鸟、猫和狗这样的种类信息，也可以是文本到图像生成和图像风格迁移中文本和图像所表述的信息。

生成器可以将简单分布（比如正态分布）转化成符合真实情况的复杂分布，即将简单分布中采样的一个样本点（比如多维正态分布中采样的一个向量）转换成一个图像。最初生成对抗网络和最初生成器没有约束信息，而在条件生成对抗网络中，条件信息将和从简单分布中采样得到的一个样本点一起作为生成器输入来进行处理，一个简单条件生成器模型结构如图3.3.4所示。

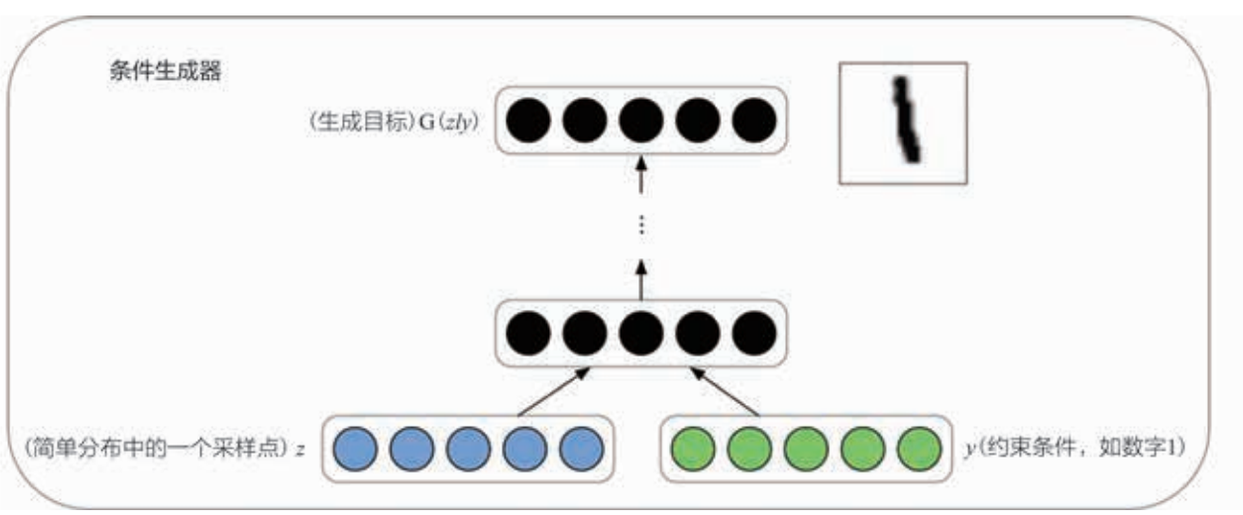


图3.3.4　一种简单的条件生成器结构

这样，生成器的输入就包含了生成过程中所需的约束信息。但是，仅仅把信息告诉生成器是不够的，因为生成器事实上还没有被约束信息所“约束”，它甚至可完全忽视这一个信息，去实现和最初生成对抗网络一样的功能。可以在判别器的设计中加入对生成数据和约束信息匹配程度的“评价”（如只有在生成器合成了既真实又符合约束规定的生成数据时，返回数值才接近1），那么生成器就会真正受到约束信息的“约束”。实现中，只需要将生成器输出图像和约束信息一起输入给判别器，就可以实现条件生成器。一种简单的条件生成对抗网络如图3.3.5所示。

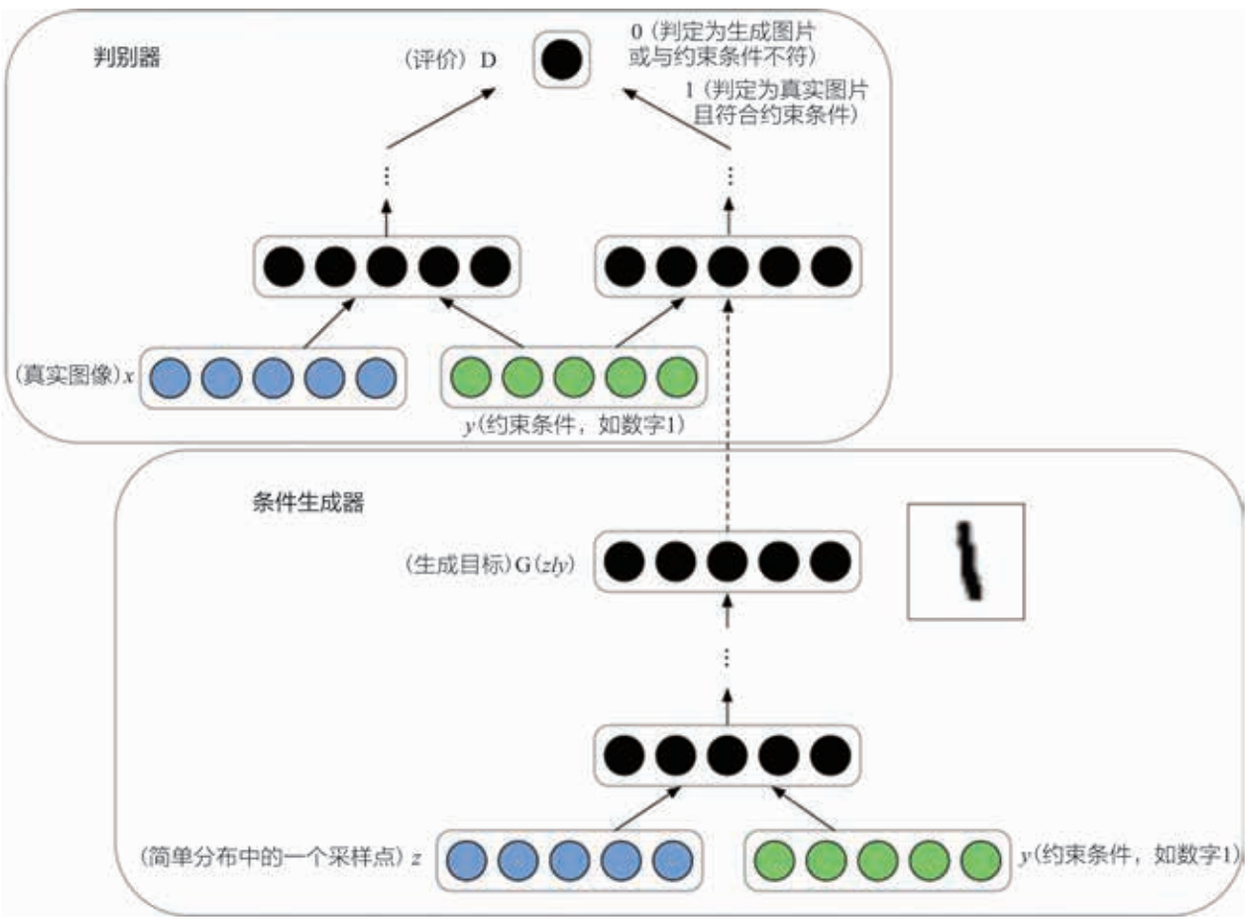


图3.3.5　一种简单的条件生成对抗网络结构

当条件约束是文本，生成目标是图像时，条件生成对抗网络所生成的例子如图3.3.6所示。

图3.3.6　约束条件为文本时图像生成的例子

当条件约束是文本和图片，生成目标是图像时，条件生成对抗网络所生成的例子如图3.3.7所示。

图3.3.7　约束条件为文本和图像时图像生成的例子

思考与练习

1. 对抗学习中，生成神经网络和判别神经网络在学习中不断发现不足，迭代式学习。请简述生成神经网络和判别神经网络的学习训练过程。

2. 如果使用人工智能的手段解决下列问题，应该用判别模型还是生成模型？请说明理由。

（1）根据已有股票价格走势，预测未来一周某只股票整体走势。

（2）根据微博讨论内容，判断用户对某一数码产品的评价。

（3）将一个中文句子分割为一个一个的词。

（4）生成某一视频所描述的内容。

（5）将中文翻译成英文。

（6）将一个人的侧面照片转变为正面照片。

巩固与提高

1. 给定两组数据 {43, 54, 67, 65, 65, 93} 和 {0.33, 0.45, 0.79, 0.56, 0.82, 0.83}，可以看出两组数据之间的差距差别很大，如第一组数据中最大值和最小值之差为 50，第二组数据中最大值和最小值之差为 0.50，两者差距为 100 倍。这样，机器学习算法在处理这两组数据过程中，会使得第二组数据被“湮没”在第一组数据中，第二组数据难以发挥作用。请利用 3.1.2 小节介绍的极大极小值归一化方法对这两组数据进行归一化处理。通过归一化操作，两组数据被规整到 [0,1] 区间，使得两组数据在后续机器学习算法中可“平等”参与计算。

2. 某城市有 5 个供应商 A，B，C，D，E，其坐标位置如下：

$$A(0,1)\quad B(0,-1)\quad C(2,0)\quad D(3,-1)\quad E(6,0)$$

为了提高货物供应效率，需要为这五个供应商建立 2 个仓储基地。请使用 K 均值（K-means）聚类算法将这五个供应商聚类为 2 组，以便将聚类质心位置作为仓储基地。

在 K 均值聚类过程中，距离函数定义为欧氏距离，聚类过程迭代两次结束。

如果给定的两个初始聚类质心分别为 $Z_1^0(0,0)$ 和 $Z_2^0(3,0)$，进行两次迭代聚类。迭代过程如下：

（1）第一次迭代

	A	B	C	D	E
到聚类质心 Z_1^0 距离	1	1	2	3.16	6
到聚类质心 Z_2^0 距离	3.16	3.16	1	1	3
所属聚类集合	Z_1^0	Z_1^0	Z_2^0	Z_2^0	Z_2^0

可见，第一次迭代后，A，B 属于同一聚类集合，C，D，E 属于同一聚类集合，接下来更新每个聚类质心。

$$Z_1^0=\frac{A+B}{2}=\left(\frac{0+0}{2},\frac{1-1}{2}\right)=(0,0)$$

$$Z_2^1=\frac{C+D+E}{3}=\left(\frac{2+3+6}{3},\frac{0-1+0}{3}\right)=\left(\frac{11}{3},-\frac{1}{3}\right)$$

（2）第二次迭代

	A	B	C	D	E
到聚类质心Z_1^1距离	1	1	2	3.16	6
到聚类质心Z_2^1距离	3.90	3.73	1.70	1.49	2.36
所属聚类集合	Z_1^1	Z_1^1	Z_2^1	Z_2^1	Z_2^1

可见，第二次迭代后，A，B属于同一聚类集合，C，D，E属于同一聚类集合，两次迭代聚类完成。

在K均值聚类过程中，如果初始聚类质心不同，那么聚类结果可能会存在差异。现在将初始聚类质心变为$Z_1^0(2,0)$和$Z_2^0(3,0)$，请计算给出两次迭代的聚类过程。并将新的聚类结果与上面的聚类结果进行比较，观察不同初始聚类质心对聚类结果的影响。

3. 滤波器的典型作用包含图像模糊（或图像平滑）以及图像去噪。除了图3.2.4中的高斯滤波器，常用的滤波器还有如图3.3.8所示的平均滤波器。在平均滤波器中，权重均为1。下面介绍两个案例：利用平均滤波器进行图像模糊操作和图像去噪操作。

1	1	1
1	1	1
1	1	1

图3.3.8　平均滤波器

（1）模糊

如图3.3.9所示，使用平均滤波器对灰度图像进行10次迭代滤波操作。可见，在滤波之前的图像中，相邻像素点之间的灰度值差别较大；经过10次滤波操作后，相邻像素点之间的灰度值差别在逐渐减少。由于相邻像素点之间的差别减少，这使得像素点之间的区别度降低，因此图像变得模糊。

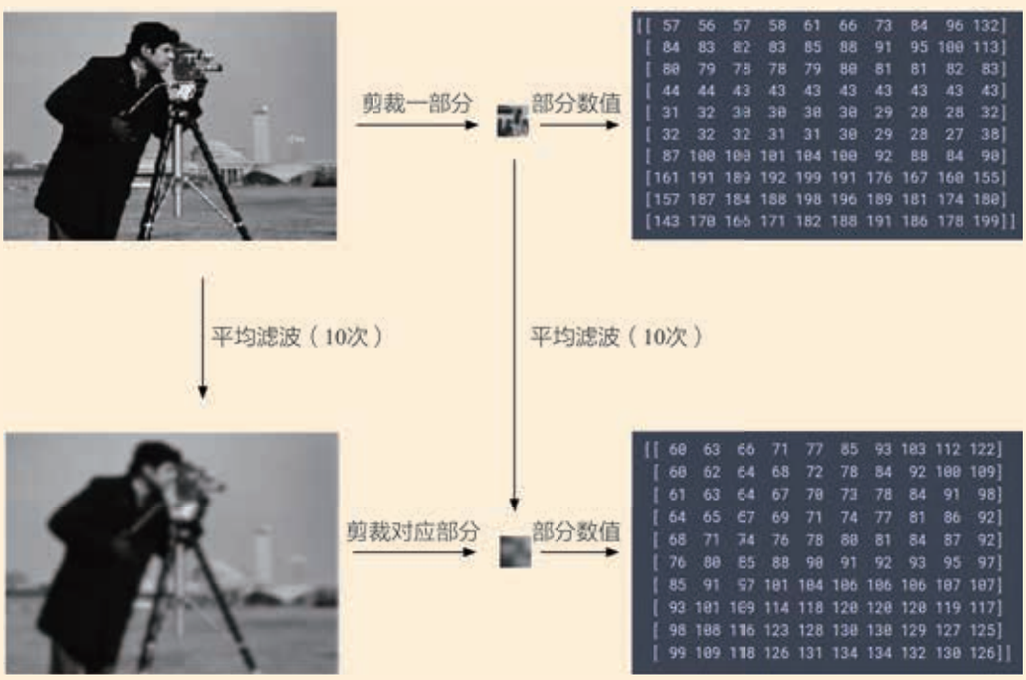


图3.3.9　图像模糊

（2）去噪

如图3.3.10所示，使用平均滤波器对原图进行一次滤波操作。可见，原始图像中的噪声点（即白色像素点）在滤波后的结果中不是特别明显，因此起到了去噪效果。

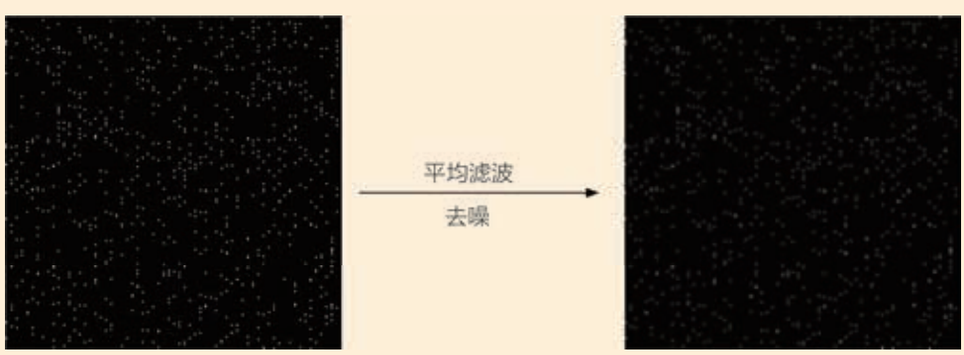


图3.3.10　图像去噪

请用平均滤波器对下面以65和85为中心的两个3×3的区域经过一次滤波，并解释你的发现（为什么平均滤波可以实现模糊和去噪）。

77	75	70	76
74	65	85	77
75	75	80	74

4. 在第二章巩固与提高第4题中，给定了三个样本（每个样本包含三个输入）和神经元的参数，可计算出神经元的输出。现给定相同的样本和参数，但是在神经元中增加一个激活函数g。激活函数g是一个符号函数，当其输入值大于或等于0时，g输出为1；否则为–1。这样神经元可完成分类任务，将输入分类为1或–1两个类别。

试计算表3.3.1中的三个样本分别属于哪一类别。

表3.3.1　三个样本

	输入1（x_1）	输入2（x_2）	输入3（x_3）
样本1	1	1	1
样本2	2	1	1
样本3	1	1	0

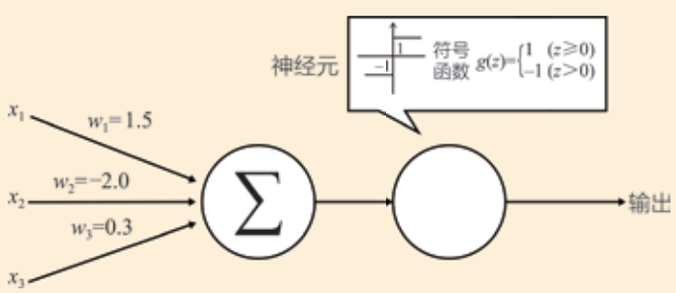


图3.3.11　感知模型示意图

注：激活函数是深度学习中神经元的一个重要功能，它实现了一种非线性映射。如本题所设计的激活函数将正数映射为1、负数映射为−1。深度学习模型可将“人脸图像”（像素点表示）识别为“人脸”（自然语言表示），这本身是一种复杂的非线性变换。深度学习中，众多神经元组合在一起，通过各自非线性映射构成了一种复杂非线性变换，使得神经网络具有强大的非线

性映射能力。

5. 1936年，艾尔弗雷德·兰登与时任美国总统的富兰克林·罗斯福同时竞选下届总统。成功预测1920年、1924年、1928年、1932年连续4届美国总统大选结果的《文学文摘》开始了其选情预测任务。在这次预测中，《文学文摘》给自己的订户邮寄了1000万份调查问卷，根据回收到的240万份回执，预测兰登将会以55比41的优势击败罗斯福赢得大选。与此同时，美国数学家、抽样调查方法创始人、民意调查的组织者乔治·盖洛普仅通过3000人的问卷调查，得出了准确得多的预测结果。盖洛普成功的原因在于根据选民的职业、年龄等科学采样，保证抽样的随机性，而《文学文摘》将采样用户锁定为其订阅客户（这些客户多集中于中上阶层），导致了样本偏差（数据分布不均匀）。可见，良好的“小数据”可战胜“含噪和偏差”的大数据，小数据可做到“以小见大”“一叶知秋”。数据质量对于机器学习预测结果影响甚大。

请设计一种数据采样方法和分类方法（如按照年龄、职业、收入等因素对人群进行分类），以便能够对某个城市居民网上购物的兴趣偏好进行统计分析。

项目挑战

实现迷你智能校园系统

自从人工智能出现以来，如何使用人工智能技术更好地服务社会，一直是人们广泛关注的话题。如今，出于安全角度考虑，各类监控系统被广泛应用于校园来保障学生的安全。但是，如何更充分地利用这些监控系统所拍摄到的图像信息，是一直困扰科技人员的问题。利用人工智能的知识和技术帮助学校进行智能化管理校园，减轻安保、后勤教职员的负担则是一个非常有意义的任务。根据已经学习的算法，设计并实现一个迷你智能校园系统。

项目任务

以小组合作的形式，自主搭建一个能够对已经截取完毕的车牌图像进行识别的神经网络模型。

过程与建议

车牌识别主要包括两个步骤，首先需要从图像中截取出汽车车牌的区域，然后对区域内的图像进行字符的识别。在本任务中，只要求对截取好的、大小统一的车牌图像加以识别。

真实车牌图像包含了用户隐私数据，并不适合私人使用，同时也难以大规模进行采集。因此，在实验中使用的是依靠程序生成的大量模拟车牌照片图像。

实验建议使用Keras来搭建模型。同学们可以查阅Keras中文文档，了解其提供了哪些不同的连接层；也可以查阅相关的文献和资料，了解效果较好的网络结构。小组成员可以尝试设计不同的模型并进行效果检验，以找出效果最好的组合。

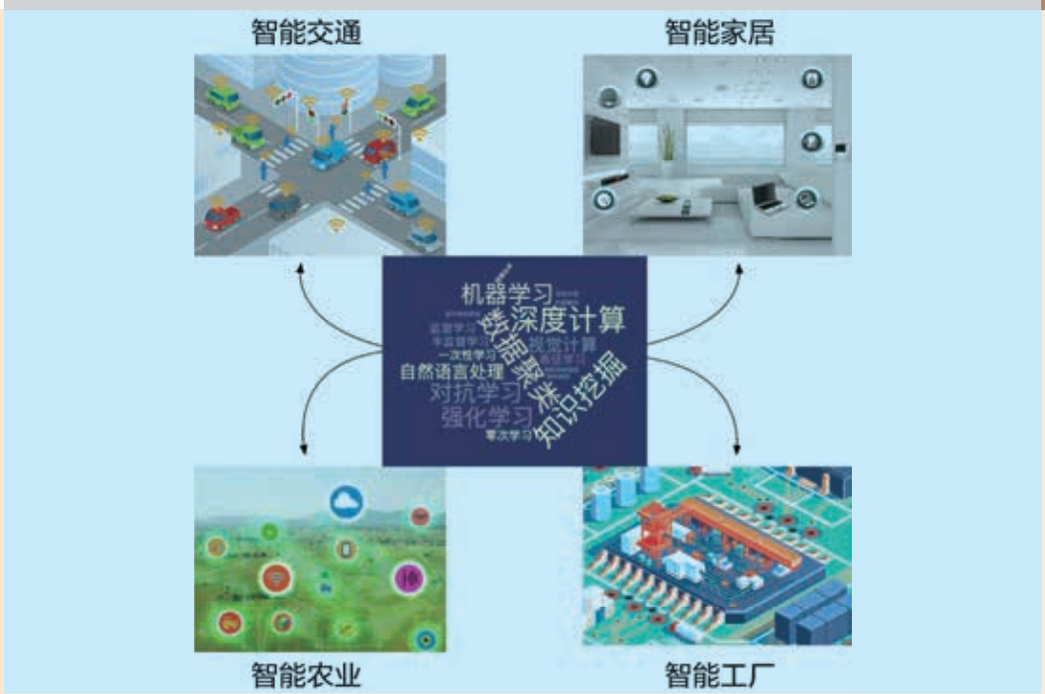
评价标准

请根据项目实施的过程、效果以及成果展示交流的结果，对自己完成项目的情况进行客观的评价，并思考后续完善的方向。把评价结果和完善方案填写在下面的表格中。

评价条目	说明	评分（1～10分）	评分主要依据阐述	后续完善方向
设计神经网络	设计的神经网络结构清晰，可以解决该分类问题			
训练和测试	在训练和测试数据集上有较为良好的表现			
展示交流	在规定时间内清晰、有条理地介绍研究成果			

第四章 4

智能之用：服务社会



人工智能在其诞生之日所立下的宏伟目标是让机器像人那样认知、思考和学习，即用计算机模拟人的智能。经过60多年的发展，人工智能的目标已变为模拟人类某一部分功能或人机共融协同发展以解决问题。

本章主要从机器翻译、人机博弈、无人系统、脑机交互四个方面介绍人工智能如何服务社会、推动智能社会的发展。机器翻译的核心是自然语言理解，这是推动自然人机交互（如语音识别）的核心技术之一；人机博弈反映了机器智能的水平和高度，一直是人工智能发展的重中之重；无人系统则是将人的感知系统（如视觉和触觉等）与认知系统（如识别和推理等）紧密结合的典范；脑机交互则为未来人类智能和机器智能自然融合打下了坚实基础。

问题与挑战

- 一个人工智能系统会模仿人类某一功能或某些功能，使其具有“类人”智能。本章介绍了机器翻译、人机博弈、无人系统和脑机交互四个应用。这四个应用分别实现了人类智能中感知、认知、推理等哪些方面的内容。

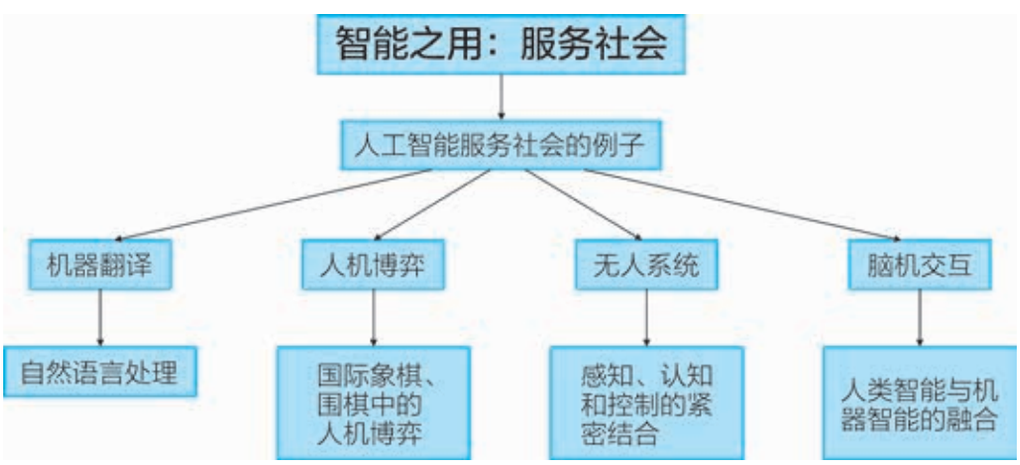
- 人工智能系统一般在接收输入数据后，对输入数据进行分析和理解，然后从已有设定中寻找与当前任务或场景最匹配的情况，进而进行决策。可见，现有人工智能系统难以处理其从未设定的任务或场景。试思考人工智能系统在未来能对从未定义的任务或场景进行分析与处理的可能途径。

- 无人驾驶车系统在相当长时间内行驶在复杂路况中还会是“有人”系统，无论人在车上还是通过网络来控制无人车。完全意义下的“无人”驾驶系统会在满足一定条件的路况环境中得到应用。思考完全意义下“无人”驾驶系统可应用范围以及有人的“无人”驾驶系统中人会发挥怎样的作用。

学习目标

1. 了解人工智能对社会生活智能化的促进作用。
2. 了解机器翻译、人机博弈、无人系统和脑机交互等人工智能应用所实现方法及其局限性。
3. 对人工智能发展给社会带来的潜在影响有正确认识。

内容总览



4.1 “智能＋X”推动社会进步

“人工智能”这一术语在达特茅斯会议上被首次提出后，人工智能经历了60多年的发展和积淀。随着互联网、大数据、云计算和物联网等技术不断发展，人工智能正引发可产生链式反应的科学突破，催生一批颠覆性技术、培育经济发展新动能、塑造新型产业体系，加速新一轮科技革命和产业变革。

以蒸汽机为代表的第一次工业革命开创了蒸汽时代，蒸汽机延伸了人的肢体，增强了人类的力量。当今人工智能与大数据、云计算和物联网等技术将耦合在一起，放大人类智力所能创造的价值，推动人类社会由数据化和网络化时代迈向智能化新的发展阶段。

当今人工智能发展的特点是应用驱动。人工智能正渗透到各行各业，不断提高实体经济发展的质量和效益，通过“智能＋X”发挥出“头雁效应”。

如图4.1.1所示，人工智能在模型、算法、算力、数据和场景等元素的综合叠加下，与大数据、物联网、互联网、5G、机器人等技术紧密结合，辅以法律法规、伦理规范和政策体系的把控，正在从推动传统行业升级、塑造新产业以及增加新体验等若干方面推动社会进步，呈现“至小有内”的内涵特性以及“至大无外”的渗透特点。

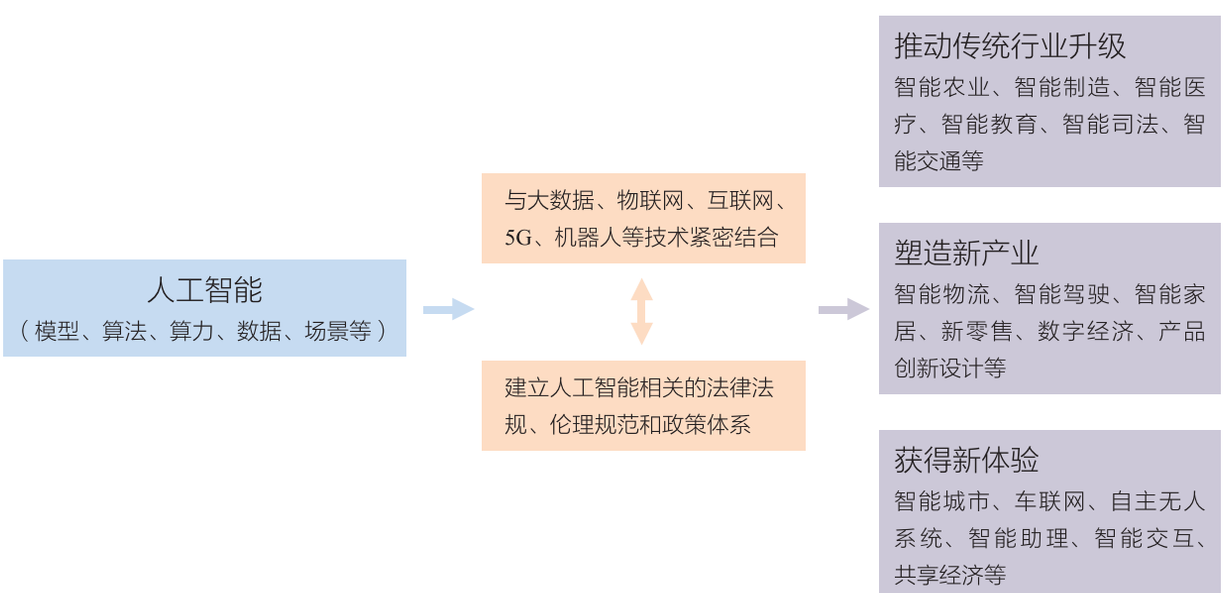


图4.1.1 “智能＋X”推动社会进步

在推动传统产业升级方面，“智能＋X”可提升传统行业的效率。例如，智能农业从种植、销售和服务等方面实现农产品的全链条管理；智能制造能够让消费者在设计阶段按照自己的偏好选择产品（如汽车）外观、设定各种性能参数，将定制产品的需求发布到云端，智能工厂接收到用户订单后进行生产任务分解并预先在虚拟工厂中进行生产线重构和

生产仿真验证，继而投入实际的生产；智能医疗系统能够辅助医生进行资料汇总查询以及为疑难病症治疗提供参考方案；智能教育系统实现个性化学习、终身学习；智能司法系统能够帮助法官分析案例；智能交通系统将路况、人流和车流以及城市功能布局等信息进行统一考虑，提高交通效率。

在塑造新产业方面，可以看到智能物流、智能驾驶、智能家居、新零售、数字经济和产品创新设计等正在或即将走入人类生活，引发经济结构重大演化，改变个人生活方式甚至社会结构，实现社会生产力的整体跃升。

在人工智能的推动下，一些新的系统或设备相继问世，给人们提供了无穷无尽的新体验。如智能城市系统合理调配城市资源，用智能化手段来提升“最多跑一次”等体验；车联网系统连接道路交通工具，更好统筹交通资源；自主无人系统勇闯生命禁区，如中国嫦娥四号探测器成功着陆在月球背面，并通过“鹊桥”中继星传回了世界第一张近距离拍摄的月背影像图；智能助理可进行导航、订票和订餐等服务；智能交互通过语音识别、情感分析和视觉理解等手段，在人类和机器之间搭建起沟通的桥梁。

本书后续章节主要从自然语言理解、竞智博弈、智能控制和人机混合智能等技术角度来介绍机器翻译、人机博弈、无人驾驶车系统和脑机接口等内容，这些智能技术已经或将会深刻影响人类生活和社会发展。

4.2 自然语言理解：机器翻译

机器翻译是以自然语言理解为核心的一门技术，其利用计算机算法将一种自然语言（源语言）转换为另一种自然语言（目标语言）。机器翻译属于自然语言处理和计算语言学范畴，具有重要的科学研究价值和实用价值。随着经济全球化及互联网的飞速发展，机器翻译技术在促进文化交流和经济发展等方面起到越来越重要的作用。

从早期词典匹配方式的机器翻译到词典匹配过程中结合语言学专家知识的机器翻译，再到基于语料库的统计机器翻译，随着计算能力的提升，互联网上丰富的多语种语料数据以及机器翻译技术逐渐走出象牙塔，开始为普通用户提供实时便捷的翻译服务。

4.2.1 基于语法规则的机器翻译

早期的机器翻译主要基于语法规则，可分为如下三种：源语言单词—目标语言单词的翻译、源语言句子结构—目标语言句子结构的翻译、源语言—中间语言—目标语言的翻译。

在源语言单词—目标语言单词的翻译中，计算机算法会把两种语言对应的单词存储在字典中，然后逐词一对一完成翻译。可以想象，在这种翻译中，很难处理一词多义或从未在字典中出现过的单词。

在源语言句子结构—目标语言句子结构的翻译中，不仅需要考虑单词与单词之间的匹配程度，而且还要考虑单词所形成的结构（如动名词短语等）之间的匹配程度。

源语言—中间语言—目标语言的翻译较好理解，如果源语言和目标语言之间无法直接翻译，可以先翻译成中间语言，然后通过中间语言实现两种语言的翻译。

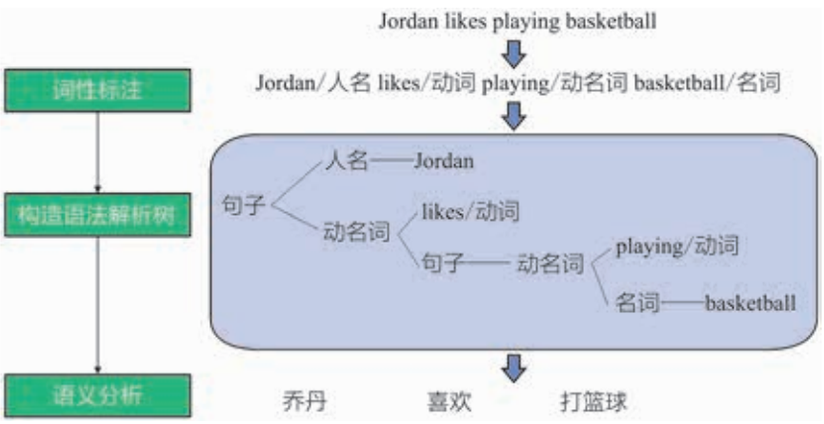


图4.2.1 基于语法分析的机器翻译示意图

智能算法经过词性标注、构造语法解析树、语义分析等过程，挖掘出句子中的结构

（如动名词短语等），实现了单词语义翻译，进而合成翻译后的中文句子。如图4.2.1所示，“Jordan likes playing basketball”被翻译成“乔丹喜欢打篮球”。

4.2.2 基于语料库的机器翻译

20世纪90年代，随着计算能力、存储容量的提升以及大量双语语料库的出现，基于语料库（corpus-based）的机器翻译技术得到了迅速发展。语料库就是经过选择和加工的大规模电子文本库，如双语单词词典、双语句子词典等。

表4.2.1 中英语料库示意

编号	源语言句子	目标语言句子
1	第一届现代奥林匹克运动会于1896年在希腊雅典举行。	The first modern Olympics were held in Athens, Greece, in 1896.
2	中国古代四大发明是指南针、火药、造纸术和印刷术。	The four great inventions from ancient China are compass, gunpowder, papermaking and printing.
3	学而不思则罔，思而不学则殆。	Learning without thought is labour lost; thought without learning is perilous.
…	…	…

表4.2.1中列出了中英文对照的语句对。一旦有了这样的语句对，用户提交待翻译的源语言语句后，机器翻译算法就在源语言句子列表中寻找最相似的句子，然后将匹配句子所对应的目标语言句子作为翻译结果，提交给用户。

以中文句子翻译为英文句子为例，在语料库中，对于一个中文句子的英文翻译，机器翻译实际上是在寻找与待翻译中文句子相似的中文句子。当然，我们可以通过很多方法来判断两个句子是否相似，比如两个句子是否包含足够多的共有单词等。

从上面可以看出，基于语料库的方法实际上是完成了“句子匹配”的任务，智能算法本身对句子内容并没有完全理解。在没有理解的情况下，进行单词到单词的机械匹配式翻译，就会出现“good good study, day day up”（好好学习，天天向上）、“people mountain people sea”（人山人海）和“土博士”（Dr. Earth）等翻译结果。

翻译本身就是一种创作过程。我国著名翻译家严复在《天演论》中提到，翻译应该遵循“信、达、雅”三条原则。他写道：“译事三难：信、达、雅。求其信已大难矣！”可以看出，在没有理解句子意思的前提下进行“硬匹配式”翻译，难以保证质量，更何况是把中国古代那些精美诗句准确地翻译成英文！

因此，机器翻译不仅依赖于机器强大的计算能力，也需要娴熟掌握语言学知识以便理解语言，才能从匹配式翻译跨越到理解式翻译。

4.3 智能模拟：人机博弈

博弈是指个人或团队在一定规则的约束下，选择各自的策略来实现预定目标（如收益等）的过程。从狭义上来讲，博弈是下棋、玩扑克牌、掷色子等具有输赢性质的游戏。广义上的博弈是对策和斗智。人机博弈则是运用计算机技术和博弈思想，使计算机像人类一样从事高度智能的博弈活动。人机博弈是机器智能、兵棋推演、智能决策系统等人工智能领域研究的重要基础，同时也被认为是人工智能领域最具挑战性的研究方向。

1952 年，英国计算机科学家克里斯托弗·斯特拉奇编写出第一个西洋跳棋程序。1956 年，IBM 的亚瑟·塞缪尔开发出了第一个能够“学习”的西洋跳棋程序，即在与人类选手对弈过程中提升程序下棋的实力，并且该程序已经可以挑战具有一定水平的业余爱好者。

4.3.1 国际象棋博弈

几十年来，计算机科学家将国际象棋视为检验人工智能的标杆。1997年5月，国际象棋冠军卡斯帕罗夫和IBM公司的“深蓝”（Deep Blue）计算机在美国纽约展开了一次令全球瞩目的人机大战，如图4.3.1所示。结果，深蓝计算机发挥出色，以2胜3平1负的总比分战胜了卡斯帕罗夫，成为首个在标准比赛时限内击败国际象棋世界冠军的计算机系统，同时也成为人工智能领域的一个里程碑。

图4.3.1　世界象棋冠军卡斯帕罗夫在与IBM深蓝计算机进行第四局对决

在象棋比赛中，一般情况下希望智能算法能够判断某一时刻棋局对整个棋局胜负会带来怎样的影响。为了达到这样的目的，就需要判断该时刻之后每个棋局对整个棋局胜负的影响，然后将这些影响累加起来。但是，这种算法的实现极为困难。为此，深蓝不去判断某个时刻棋局对整个棋局胜负的影响，而是直接对任意一个棋局进行打分，来评估该棋局的优劣程度。在具体实现中，深蓝定义了8000种特征来判断某个棋局的优劣。比如：对于每一个竖线，都会有一个比特位来表示该竖线上是否有敌人的“兵”；用两个比特位表

示己方“车”超越己方“兵”的数量（最大为2）等。8000种特征中的每个特征赋予了一定权重，并且这些权重在对弈过程中可调整，使得棋局优劣判断可因局势而灵活变换。

为了选择一个最佳局面来应对当前局势，深蓝采用了一种称为“最小最大”（Min-Max）的原则。“最小最大”的含义是使己方收益最大、对手却使自方收益最小。在如图4.3.2所示的游戏树中，A代表己方所在棋面，B和C代表对手所在棋面，D，E，F，G表示所对应棋面能够带来的收益（比如，棋面E的收益是−3，棋面G的收益是2）。

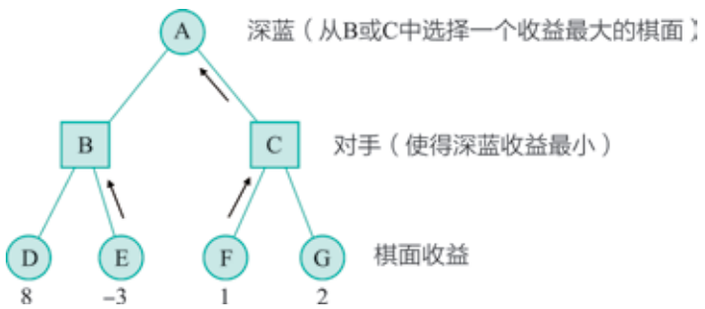


图4.3.2　Min-Max问题示例

假设深蓝位于棋面A，现在要从B或C中选择一个棋面。由于B和C是对手所在棋面，对手会尽力选择带给深蓝最小收益的棋面，因此B会选择棋面E（带来−3的收益）、C会选择棋面F（带来1的收益）。最终，深蓝为了使得自己的收益最大，会选择C（因为C选择了棋面F）。这样，深蓝获得的收益为1。

基于上述“最小最大”原则，深蓝要做的就是从当前棋面出发，对当前棋面之后的可能棋面尽可能往前搜索。由于象棋棋面众多，尽管深蓝平均每秒能够对众多棋面进行判断评估，还是无法在规定时间内通过“穷举”计算得到一个最优棋面，因此还需要一种高效的棋局搜索算法。为此，深蓝采用了Alpha-Beta剪枝搜索算法。简单来说，Alpha-Beta剪枝算法是一种启发式算法，它能够“剪掉”不需要搜索的棋面而提高搜索效率。比如，某一棋面后自己的“王”将死，就不需要以对手“放自己一条生路”为前提，继续往后搜索。平均来看，深蓝可以在3分钟内通过剪枝搜索算法将12.2步以内的所有棋面搜索完毕。事实上，除了上述的行棋策略外，深蓝还会参考已有的棋谱。这些棋谱由4000个开局棋谱、70000个中盘棋谱和众多的残局棋谱构成，其中，残局棋谱指5个或少于5个棋子的局面，或带有一对“兵”的6个棋子。通过对已有的棋谱库的搜索，深蓝倾向于选择那些在棋谱中出现率更高的走法。

4.3.2　围棋博弈

一般认为，计算机要在围棋中取胜比在国际象棋等游戏中取胜要困难得多，因为围棋的下棋点极多，而且每次落子对情势的好坏飘忽不定，诸如暴力搜索法、Alpha-Beta剪枝搜索、启发式搜索的传统人工智能方法在围棋中很难奏效，其原因在于围棋可搜索的范围更大、更复杂。

围棋有19×19个方格，如果每个方格上可放一个白棋或一个黑棋，那么理论上最多有3^{361}种棋子摆法，很难想象智能算法将所有棋子摆法都“查看”一遍，再选择一个应对当前局势的最优摆法来“出招”。

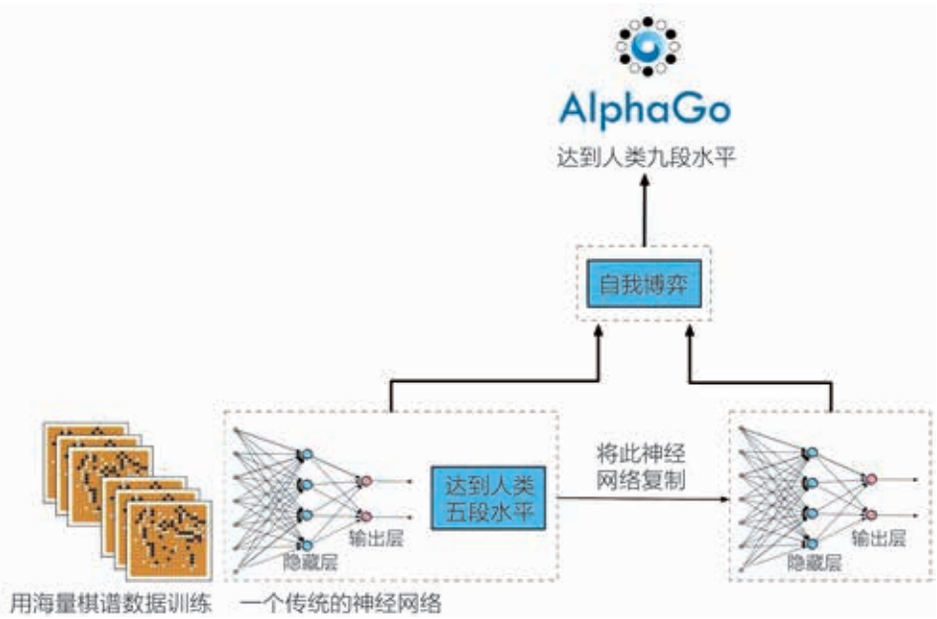


图4.3.3　AlphaGo算法架构图

AlphaGo是第一个击败人类围棋职业选手的计算机程序，它由谷歌旗下的DeepMind公司开发。AlphaGo主要包含三种算法构成（如图4.3.3所示）。

算法1：深度学习。DeepMind公司收集了成千上万个人类选手的棋局，“训练”了一个深度神经网络（卷积神经网络）来学习下棋。这个深度神经网络由于看过了海量棋局，并从海量棋局中“悟出”赢棋的模式，于是它就会下围棋了。但是，尽管该围棋程序看过了海量九段选手对决的棋局，它的围棋水平只有五段。

算法2：自我博弈。由于DeepMind公司已经成功“训练”了一个具有五段水平的围棋程序，因此将会下棋的围棋程序复制一份，通过强化学习机制让两个会下棋的围棋程序自我博弈。这两个围棋程序各自均想战胜对方，因此在对弈中快速提升能力，达到了九段水平，进而战胜了人类棋手。

算法3：蒙特卡洛树搜索。前面讲过，围棋程序为了应对当前局势，需要从众多棋子摆法中选择一种走法。若采取穷举式方法，则难以想象其所需时间。AlphaGo采用了一种称为蒙特卡洛树搜索的方法来选择一种最佳的下棋棋法。蒙特卡洛树搜索是一种采样式搜索方法，可以在没有遍历所有下棋棋法的前提下，尽可能找到应对当前局势的“好招”。

2017年10月18日，DeepMind团队公布了一种称为“AlphaGo Zero”的智能围棋程序。在前期研制的围棋程序中，需要从成千上万个人类棋手的棋局中“训练”一个会下棋的围棋程序，然后把这个会下棋的围棋程序自我拷贝一份，让两个围棋程序自我博弈，提升能力。AlphaGo Zero没有利用任何人类对弈数据，仅依靠自我对弈进行学习。经过3天的自我训练，AlphaGo Zero就打败了旧版AlphaGo，战绩是100∶0。经过40天学习，AlphaGo

Zero超过了之前所有版本的水平。

“尧造围棋，丹朱善之。”这是说尧帝发明了围棋，来教化训练其儿子丹朱。AlphaGo Zero用40天时间走完了人类4000年的围棋历史，其能力的确让人瞠目结舌。

但是，我们必须知道，围棋是一种“规则完备、约束有限、目标明确”的竞智活动。即，其下法有限、区分胜负的目标明确（数子），因此，围棋也被称为“完全信息条件下博弈”。在这种情况下，机器凭借其计算能力战胜人类选手，也是可以理解的。

与AlphaGo在已知规则下的博弈情况不同，现实社会中诸多行为决策（如新经济运行、环境变化、产业布局、网络空间安全等）是非完全信息的博弈，即在信息未能全面掌握的条件下进行推理和决策，如扑克和游戏等，这是目前人机对弈的热点问题。

拓展链接

战场博弈

东汉马融在《围棋赋》中说：“三尺之局兮，为战斗场。”可见围棋可作为模拟战场决策的博弈。不过，在真实战场对弈中，博弈双方依赖的环境并非规则方格，而是由地形、目标、参战力、武器等方面组成的复杂空间。战场博弈的搜索空间不再是简单的有限空间，而是更为庞大的搜索空间，因此是一个无限、多目标、多边界条件的优化问题。

思考与练习

棋类游戏规则和输赢判断都非常明确，可以被高度形式化和快速计算，在这种条件下计算机“技高一筹”。AlphaGo Zero没有用到任何人类选手的数据进行学习，而是依靠两个明了围棋规则的计算机程序相互对弈、相互提高来训练生成。很多人说AlphaGo Zero是无师自通，你如何看待这个观点？

4.4 智能控制：无人驾驶车系统

无人驾驶车是人工智能在社会生活中的一个重要应用。一个较为完整的无人驾驶车系统包含了雷达、激光、全球定位系统（GPS）、里程计和计算机视觉等感知/认知子系统，能够完成路线规划导航、障碍物识别和路标识别等多项任务，从而实现“无人”与“自动驾驶”的最终目标。

4.4.1 自动驾驶等级分类

按照国际自动机工程师学会（SAE）对无人车“自动化”程度的定义，可以将无人车分为六大类别，如图4.4.1所示。

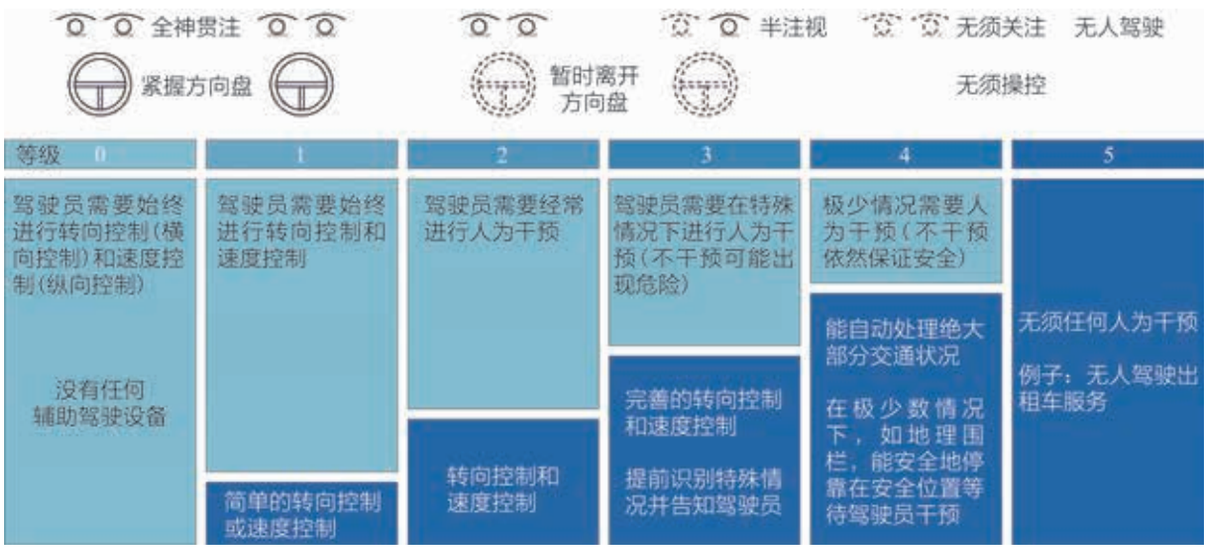


图4.4.1　国际自动机工程师学会自动驾驶分级（从左到右，驾驶员对车辆的控制程度越来越低）

- 等级0：汽车智能化系统可能会对若干危险发出警告（如倒车雷达报警或油量存量报警），但智能算法并未能实现车辆的持续控制。
- 等级1：驾驶员和自动化系统共享车辆的控制。如在自适应巡航控制（ACC）系统协助下，驾驶员来控制车辆的转向、系统来控制车速；在停车辅助系统协助下，系统来控制车辆的转向、驾驶员来控制车速。
- 等级2：自动化系统完全控制车辆（包括对加速、刹车和转向的控制）。但是，驾驶员需要随车来实时监控驾驶，并且驾驶员要做好随时干预智能系统的准备。
- 等级3：驾驶员可充分相信自动驾驶系统，驾驶员的注意力可暂时从当前驾驶任务转移到发短信、看电影等。当车辆发出干预请求时，驾驶员需要在车辆制造商指定的有限时间内进行干预（不及时干预可能发生危险）。目前，若干无人车实现了等级3形式的

自动驾驶，但只限于在特定场景（如无交通灯、相对封闭的高速公路以及工厂货品储藏间等）。

- 等级4：在等级4自动驾驶中，不需要驾驶员注意就能保证驾驶安全，驾驶员可以睡眠或暂时离开驾驶座位。在等级4情况下，仅在特殊地域（地理围栏）或特殊情况（例如交通拥堵）下需要驾驶员干预（但是，即使驾驶员不干预，仍可保证驾驶员安全）。
- 等级5：不需要任何的人为干预，这是真正意义上的“无人”驾驶。一个典型的未来畅想是无人驾驶出租车服务或无人机“快递小哥”。

4.4.2 自动驾驶系统构成

等级3及以上的自动驾驶系统通常被认为主要由系统来观察驾驶外部环境并对观测结果做出反应。一般情况下，这样的系统通常包含如图4.4.2所示的几个部分：

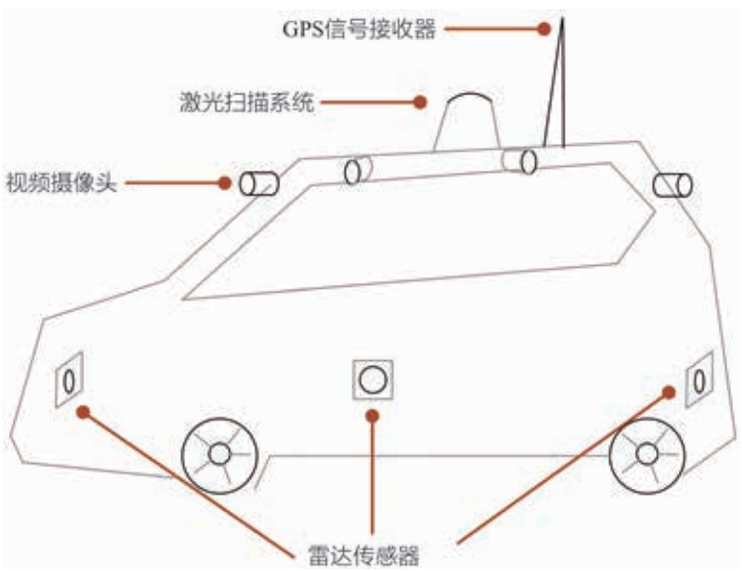


图4.4.2 自动驾驶系统示意图

- 激光扫描系统：通常被安置在汽车顶端，能够对环境进行360°的三维扫描，主要用于识别道路、辨析障碍等。
- 视频摄像头：帮助汽车识别周围环境中的视觉信息（如停止的汽车、路标路障等），跟踪行人、其他车辆的运行情况等。它还负责识别交通路灯、阅读路标。
- GPS信号接收器：依靠GPS对汽车进行精准定位，在车载地图的帮助下，无人驾驶汽车能实现最优路径规划。
- 雷达传感器：自动驾驶汽车的前后和侧面都安装了雷达系统，能够配合其他系统的检测信号，一起实现汽车的自动变道和自动泊车等。
- 车联网系统：车与车、车与路、车与人、车与云等系统间的交互称为车联网。一个安全的车联网能够帮助车辆实现信息共享，从多个信息源中获得车辆、道路、环境等信息，也可对信息进行加工、计算和发布。此外，还能提供专业的多媒体与移动互联网应用

服务。

- 智能算法模块控制系统：如果说上述子系统是无人车的眼睛和耳朵，那么控制系统就是无人车的大脑。控制系统要对上述所有子系统收集的信息数据进行分析和加工。控制系统也对各个智能算法子系统进行监督和控制，以保障整体系统的有效、稳定和安全。

4.4.3 自动驾驶技术的相关应用

在货物仓储工厂，为提高货物的分拣、搬运效率，许多公司都应用了如图4.4.3所示的货物搬运机器人，仓库中的环境与道路交通类似：多个自动行驶的机器人，每个机器人有不同的预定目标，机器人与机器人、机器人与工作人员之间都可能会相互影响。因此物流机器人需要完成包括路线规划、避障、自动泊车充电等任务，使用了许多自动化驾驶中涉及的技术：计算机视觉、深度感知、机器学习、物体检测等。

图4.4.3 仓储货物间的货物搬运机器人

除在仓储领域外，家居行业的自动扫地机器人、救灾中的营救机器人，都会用到与自动驾驶技术相同或相关的技术。

思考与练习

1. 驾驶员在等级3或等级4情况下的“无人”车中，可暂时或较长时间将注意力从驾驶任务转移开来，请写出若干在等级3和等级4情况下需要驾驶员对汽车自动系统进行干预的情况。

2. 请仔细思考，举出一个车联网系统中，车、人、路等相互协同，改变传统交通的例子。

4.5 混合智能：脑机接口

科幻小说和电影当中以外力来控制大脑的场景，一直是科学家试图解开大脑思维之谜的梦想。这一新兴领域统称为脑机接口（brain-computer interface，简称BCI）。脑机接口有时也称作“大脑端口”或者“脑机融合感知”，它是在人或动物脑（或者脑细胞的培养物）与外部设备间建立的直接连接通路。我们可以将它定义为一个不需要由周围神经和肌肉就能让大脑与外界沟通的系统。

脑机接口有三个重要组成部分，我们可以归类为三个大的方向性的问题：如何侦测脑部活动？如何从脑波信号中判读人的意志？如何用意念来控制机器或计算机？即，从大脑采集信号→对所采集的信号进行分析解码→将需要反馈的信息通过再编码回送给大脑以实现运动。

这三个问题就是脑机接口领域研究的核心。因此，现代的脑机接口系统由测量脑波信号、提取信号特征、转换成机器指令这三个元素构成，如图4.5.1所示。

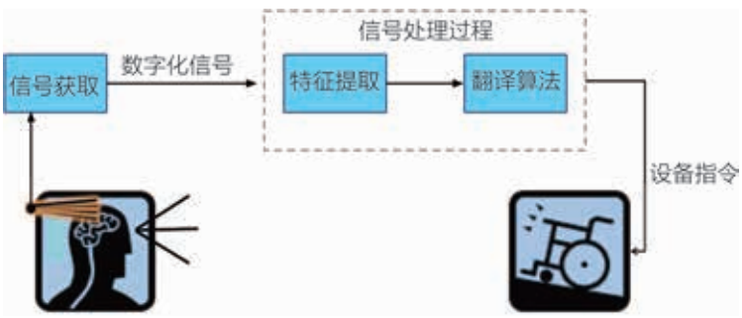


图4.5.1　脑机接口系统流程图

4.5.1　脑机接口的发展历史

大脑思维活动的产生仍然是科学上的未解之谜，最早期的脑科学仅通过解剖脑伤病人这一途径来研究大脑。1924年德国医生汉斯·贝格尔首次从人的头皮上记录到人脑发出的微弱电磁波，并发现脑波出现的振荡信号与脑部疾病相关，由此开启了探索大脑奥秘的新征程。20世纪70年代，不少科学家开始尝试用脑波信号作为与机器沟通的管道。其中最为成功的是当时UCLA BCI实验室团队实现了通过利用视觉刺激反应来控制游标走迷宫的脑机接口系统，并在1973年所发表的论文当中提出了“脑机接口”这一学术用语。

脑机接口研究中获取人脑电波信号的另外一种方式采用侵入式电极。神经科学家很早就开始以侵入式电极来研究神经细胞的电化学行为，但一直到20世纪90年代，美国犹他大学的理查德·诺曼教授开发了多电极的Utah Array，并展示通过对近一百个电极信号进行分析来完成脑机接口控制的一些简单任务。美国杜克大学医学院神经生物学教授米格

尔·尼科莱利斯等人发展了多电极测量及信号处理技术，这种以侵入式电极信号分析为手段的脑机接口广受关注。大脑控制的仿生外骨骼可视为身体的延伸，尼科莱利斯在神经可塑性领域的研究表明，大脑能够重塑自己，以适应外部环境的变化，并且融入新的元素。

21世纪开始，脑机接口以极快的速度发展，已经能够帮助失聪患者恢复听力、帮助帕金森病患者缓解症状、帮助瘫痪者实现若干活动。

如布朗大学研发的Brain Gate系统将Utah Array电极阵列植入瘫痪病人的运动皮层。当病人想要抬手或做其他活动时，电极便可检测到被这些意图激活的神经元，然后将信号传导给机器，让其完成对应指令。在Brain Gate的帮助下，一个瘫痪近15年的女士通过机械手臂完成喝咖啡的动作；2017年斯坦福大学的一位电气工程教授和一位神经外科教授让三名受试瘫痪者通过脑机接口装置Brain Gate控制屏幕上的光标，在计算机屏幕上输入了他们想说的话，其中一名患者可以平均每分钟输入39个字母，约合8个单词，这是世界上首次通过脑机接口实现如此快速、精准的字符输入。

拓展链接

神经可塑性

以前的科学研究结果表明，大脑发育到青春期后期和成年早期就结束了，大脑在成年之后就基本定型，然后随着脑细胞衰退就开始走下坡路了。现在科学发现，大脑在成年之后依旧保留着巨大的变化潜力。这种能力叫作“神经可塑性”，神经可塑性研究表明，如果一个神经元死亡了，其他的神经元将会接管它的功能。同时，我们可通过体育锻炼对我们的大脑进行训练，以增强现有的突触并创造新的突触，所谓“用进废退”。正如现代神经科学家、1906年诺贝尔生理学或医学奖得主圣地亚哥·拉蒙-卡哈尔曾说过的话：每个人都可以是他自己大脑的雕塑家。

4.5.2 脑波信号的原理与测量

大脑由百亿个神经细胞（又称神经元）组成，每个神经元又与成百上千个其他神经元连接。这些细胞通过放电的方式传递信号，并释放神经传导物质刺激其所相连的下一个神经细胞，继续传递其所接收的信息。这一复杂的神经回路构成了我们走路、记忆、思考的神经基础。

脑电信号就是大脑神经元的电活动。依据电极的侵入程度和测得信号的范围，由内部至外部、由小范围至大范围，大致可分为皮质内的脑机接口（Intracortical recordings）、皮质脑电图（Electrocorticography，即ECoG）、脑电波记录（Electroencephalography，即EEG）三种。按照观测脑电信号过程中对人体造成伤害程度的不同，可分为非侵入式（noninvasive）、侵入式（invasive）和半侵入式。

1. 脑机接口

如果打开头骨、穿过脑膜、把电极插入脑中，便可以检测到数个神经细胞此起彼落的怒吼（即神经脉冲，spikes），如图4.5.2所示。若将信号接到麦克风，即可听到清晰的神经元放电声音，亦称为脉冲序列（spike train）。局部场信号（local field potentials，简称LFP）是其中的一种皮质内信号。

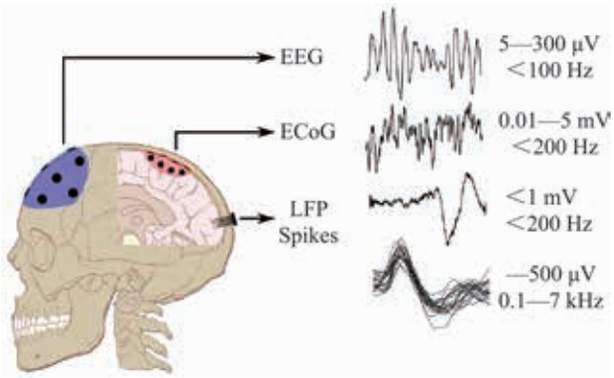


图4.5.2　三种脑信号示意图

此方法常被神经科学家用来研究大脑某些特定的功能，因为普遍认为脉冲序列包含了脑活动的所有信息，是最直接测量脑部活动的方法（相较于ECoG及EEG），所提取的信号质量也很高。但是也有两个主要的问题：第一，侵入式的电极可能会造成脑损伤，所以目前研究皆以昆虫、老鼠、猴子为主。第二，身体对外来物会产生免疫排斥反应，随着时间的推移，所记录到的神经元数目会逐渐减少，记录到的信号会逐渐变弱，从而影响到脑机接口的性能。

2. 皮质脑电图和脑电波记录

ECoG和EEG的信号就是我们俗称的脑电波，两者的信号产生原理几乎相同。神经元活化产生的动作电位，透过轴突传递到突触之后，释放的神经传导物质会引发接棒的神经元产生突触后电位，进一步刺激或抑制该神经元产生动作电位。突触后电位使神经元产生均匀不一的电荷分布，导致电荷流动。当一群神经细胞同步被活化，就会产生足够大的电流信号，形成相应电磁场，从脑细胞间质、脑膜、脑组织液、头盖骨，一层层穿透至头皮，最终成为我们测量到的脑波信号。

ECoG与EEG的差异在于，ECoG测量是将患者头盖骨打开，把电极平放在大脑皮层上与硬脑膜下的半侵入式方法，EEG则是在头皮上非侵入式直接测量脑波。ECoG好处是减少来自脑膜、脑组织液、头盖骨等介质的干扰信号。EEG的优势则是不需要手术，较为便利和成本较低。

ECoG和EEG虽然空间准确度不高，但电极可涵盖整个脑，观测范围比皮质内的脑机接口（仅能观测数厘米）要广泛得多。因此ECoG常被应用在临床上以协助癫痫病患定位不规则活动的脑区，以利于开刀切除。EEG则用来发现与许多认知功能相关的脑区（如专注程度、情绪等）以及神经疾病（如自闭症、精神分裂症等）。

4.5.3　脑波信号的特征提取与处理

取得脉冲序列或脑波信号之后，接下来的问题是：“它告诉了我们什么？”如何从复

杂的、大量的信号里提取有用的信息，进而判读人的意念，这是脑机接口领域最有挑战性且最热门的议题之一。

测量得到的脑波是数量庞大的神经元电信号与各种噪声的杂合结果，还无法精确地破译其中的信息，也还不存在完整的理论解释脑波信号及神经网络、脑功能的关系。所以目前脑电波分析方法仍采用数据驱动（Data-driven）手段。神经科学家、信号处理专家、统计学家，尝试通过各种分析方法，寻找大脑在执行不同任务或处于不同状态下的脑波特征。虽然目前尚无法解析脑波中全部的信息，但只要能够判读大脑状态或意图所隐藏的若干信息，就能通过脑机接口界面实现大脑与外界的简易沟通。

若能提取显著的脑波特征，将其转换为指令来控制电脑或机器，就实现了意念控制机器。这样，解读人的所有意念是否并非难事了呢？

脑机接口实际上还存在两大问题：第一，实际所关心的信号会与诸如眨眼等肌肉电信号（EMG）和周围电磁波（如手机、计算机）干扰信号等混合在一起，观测所得的信号往往比真正脑波庞大得多；第二，复杂的脑活动不是仅靠单一特征就能预测，许多脑功能（尤其是高阶认知功能如推理、情感等）仍找不到指标性的特征来度量。

那么该如何从混合于一体的庞大信号中萃取真正关切的脑波信号呢？面对复杂未知的脑活动，我们还有什么办法去判读蕴藏其内的秘密呢？这两个问题对应了科学家正努力的两大方向：一是信号精准分析，包含噪声去除等；二是开发新的分析方法，例如采用机器学习里的算法技术。

传统的信号处理方法诸如滤波器即能有效滤除如交流电60Hz电磁波或直流电的干扰。但肌电信号的干扰一直困扰着科学家，它往往会降低脑机界面判读脑波特征的准确率。一种称为独立成分分析（Independent Component Analysis, 简称ICA）的方法被用于解决这一问题。

ICA假设脑波是由许多独立的脑区发出信号混合而成，用此模型可解译脑波内含的独立成分，如分离出眨眼、眼动等肌肉电信号。

机器学习与脑波分析有什么关系呢？机器学习的目的，是从大数据中找出数据所蕴含的模式和规律。如从数据中预测用户的行为和喜好、从脑波信号中来判读大脑意念。

近十年来，机器学习方法的进步促进了神经科学大数据的发展。通过机器学习，从大量看似无意义的脑波信号中，判断复杂的脑活动（如精细的运动控制），大幅提升脑机界面判读的准确率。回归分析、人工神经网络等在之前的章节中学习过的方法也都能在脑机接口领域应用，从而促进脑机接口蓬勃发展。

思考与练习

与脑机接口有关的电影数不胜数，且电影拍摄时期相对应的技术往往还没有出现。你能发挥想象力来畅想一项新的潜在技术或应用吗？

4.6 人工智能发展对社会的潜在影响

人工智能正在推动人类社会由信息化时代迈向智能化时代，在这场变革中，我们应对人工智能给社会带来的潜在影响建立正确、客观的认识。

本章介绍了人工智能在机器翻译、人机博弈、无人系统和脑机交互等方面的技术。这些技术像一把双刃剑，一方面给生产、生活带来了极大便利，另一方面也产生了消极影响，如使人们变得过于依赖机器智能，动手能力衰退等。

回顾历史，在英国工业革命时期，机器生产逐渐替代手工劳动使大批手工业者失业，失业工人曾联合起来捣毁机器进行抗争，这一事件在历史上称为“卢德运动”（Luddism）。从那以后，“卢德运动”成为抵制新技术、对抗社会发展的倒退行为的代名词。今天，为了避免重蹈覆辙，我们对于人工智能发展会取代人类部分工作（特别是那些标准化、重复性的工作，如电梯操作员）这一问题，要持有正确态度：人工智能是人类发明出来的一种工具，它会极大提升人类的工作效率。人工智能会促进传统产业的升级换代，提高生产效率，对就业结构产生影响，这虽然会直接导致某些工种消失和部分劳动力失业，但人工智能发展会塑造新的劳动模式、带来新的工作机会，人们应当积极提升自己的技能，适应技术变革，胜任在技术变革中产生的新工作。只要对人工智能发展所带来的潜在负面影响保持清醒认识，人类就会掌握好人工智能这把利剑，为人类带来莫大福祉。

思考与练习

在你身边有哪些人工智能与人类智能互补、合作的例子？设想一下，在未来的十年内，人工智能将会在哪些领域取得长足进步，并更好地服务人类？

巩固与提高

1. 用现有的机器翻译软件（如有道词典等），分别尝试三句英文到中文、中文到英文的翻译。查看翻译结果，如果觉得翻译得不好，请指出其中的不足之处。在实际操作中，可以尝试对“寻寻觅觅，冷冷清清，凄凄惨惨戚戚”等诗句进行翻译，可以从“信、达、雅”角度来评价翻译结果的质量。

2. 回顾深蓝的策略——最小化对手可能拿到的最高分。请指出在如图4.6.1所示的局面中，深蓝应该选择B，C，D之中的哪种局面。

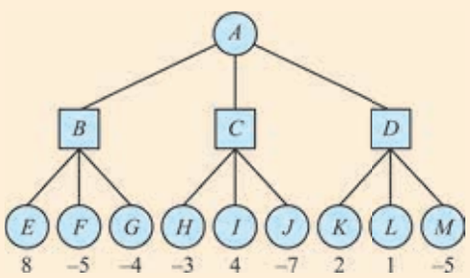


图4.6.1

3. 请简述自动驾驶系统的构成及各个模块的作用。

4. 脑机接口可以帮助失聪患者恢复听力、帮助帕金森病患者缓解症状、帮助瘫痪者实现若干活动。请查找资料，想一想脑机接口还可能有哪些应用。

项目挑战

创建智能家庭港湾

家是我们忙碌一天后休息和放松的港湾。现在，已经学习了许多人工智能技术和人工智能应用的你，希望设计一个智能音箱，它通过语音控制，既能够调节灯光亮度、空调温度、给特定的人发送信息、预定出租车、查询天气等，为父母换来更多的休息时间，也能够播放音乐、新闻、音频书，帮助父母放松心情。希望你能够通过这次人机交互实践，真正掌握“智能之用”，服务家庭、服务社会。

项目任务

设计一个由语音控制的多功能智能音箱，帮助人自动完成琐碎和重复的工作。具体要求如下：

1. 你需要尽可能地了解和设想智能音箱可以实现的功能。

2. 你需要剖析基本技能，这些技能可以使用哪些人工智能技术或方法实现。

3. 你需要画出实现一个具体功能（如播放音乐），从语音输出到最后功能实现所涉及的所有技术的流程图或技术路线。

4. 你需要利用人工智能方法尝试实现智能音箱最为核心的基本技能——语音到文字的转换。

过程与建议

1. 确定基本技能

设想在家庭生活各种琐碎的小事，这些小事可以使用哪些功能去实现，这些功能又可以通过哪些基本技能来实现。你可以尝试使用头脑风暴的方式来完成。（注：基本技能与具体应用功能无关，基本技能的组合可以实现具体应用功能）

2. 剖析基本技能

归纳整理上一步得到的基本技能，对于每一种基本技能，考虑可以使用哪一种人工智能技术或方法来实现。（注：你可以从这些基本技能的输入输出着手，思考使用的人工智能技术或方法）

3. 构建技术路线

从第一步设想的应用功能中任意挑出一个（比如播放音乐），考虑从最初的输入（语

音）到最后的输出或行为（播放语音中指定或描述的音乐）中可能涉及的各种技术和方法，以流程图的形式展示数据或逻辑走向。

4. 核心技能实现

使用智能音箱，无论完成哪一个应用功能，都需要一个基本技能：从语音到文字的识别转换，你需要进行实践，使用之前所学的人工智能技术或方法来实现这项基本技能，类似于之前的实践任务，你可以：

（1）收集数据。你可以搜索和使用公开的语音到文本的数据集，也可以自行搜集（录下语音，记录下对应文本）。

（2）数据预处理。对于语音数据，你可以借助一些语音特征抽取工具（如pyAudio等）来从语音中抽取一些基于频率的特征，如音谱（Spectrum）等；对于文本数据，可以使用在机器翻译中学习到的预处理方法（如分词、构建词表等）来对文本进行预处理。

（3）设计和构建模型。考虑之前学习过的人工智能算法，选择你认为最合适的方法构建模型。

（4）编码和训练。用代码实现你设计的算法，并在CPU（或GPU）上对模型进行训练。

（5）测试。输入任意语音，测试模型的效果。

5. 撰写项目报告

基于以上工作，撰写分析报告，其中应包括如下内容：

（1）研究背景与目标。

（2）智能音箱技能与功能的描述。

（3）实现特定功能的技术路线图。

（4）核心技能实践效果及评价。

（5）结论与建议。

评价标准

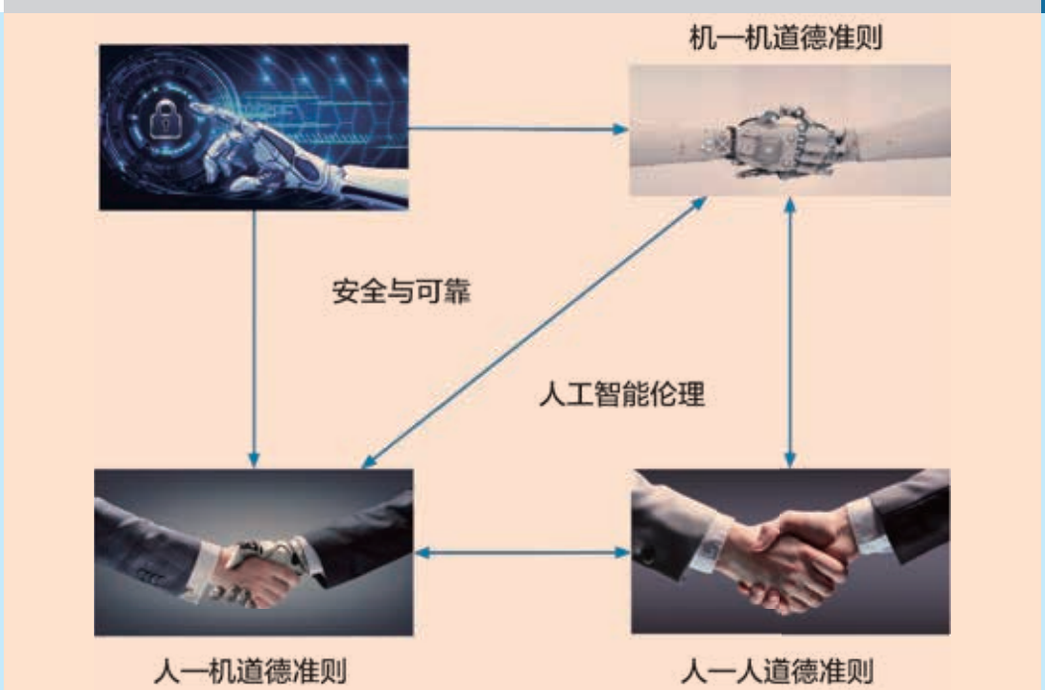
请根据项目实施的过程、效果以及成果展示交流的结果，对自己完成项目的情况进行客观的评价，并思考后续完善的方向。把评价结果和完善方案填写在下面的表格中。

评价条目	说明	评分（1～10分）	评分主要依据阐述	后续完善方向
项目理解	在项目开展和分析报告中，展现出对项目目标、内容与任务的正确理解			

评价条目	说明	评分（1～10分）	评分主要依据阐述	后续完善方向
小组协作	小组分工合理，协作紧密，合作有成效			
功能完整性	功能设想的完整性，能满足大部分日常使用的需要			
技能完整性	技能设想的完整性，能通过相互组合实现大部分功能			
数据采集	数据来源可靠、数据获取方法合理			
模型构建	模型的有效性和可行性			
实践效果	实践结果的效果评价			
展示交流	分析正确，展示经过精心准备，表达清晰、便于受众理解			

第五章 5

智能之基：伦理与安全



虽然人工智能在近几年取得了可喜的技术进步，但其实仍处在“婴幼期”。正如年幼孩子会模仿社会中的成人行为一样，人工智能也正在向人类学习。因此，作为人工智能的“父母”，人类必须确保把价值观和道德观贯穿于人工智能的产品和服务。在一般意义上，伦理关注人与人之间的道德规范和准则，人工智能伦理更多关注的是人—机、机—机以及人—机共融所形成的社会形态应该遵守的道德准则，因此建立人工智能伦理是保障人工智能深入人类社会生活需要考虑的重要问题。

人类不仅要从法律法规和道德准则方面来保障人工智能的发展，还要通过技术手段来保障人工智能的安全性。这包括如下两个方面：一是人工智能算法运行结果要与算法设计期望目标保持一致，即程序是“正确的”。为了确保程序的“正确性”，可对程序所处理的各种情况进行“穷举式测试”（往往不可能）或者进行毕其功于一役的“形式化证明”（往往很困难）。二是人工智能系统应该具有可靠性。在现实生活中，人工智能系统会被黑客攻击，如在无人车系统中，黑客可能会远程操控无人车或在识别样本中添加噪声，使得人工智能系统在行驶过程中错误辨识目标而造成严重影响（如将行人辨识为建筑物）。因此人工智能系统要做好应对恶意攻击的准备，也就是说智能算法要具有鲁棒性（即算法的可靠性）。

问题与挑战

- 比尔·盖茨曾经说过：“微软不是一家软件开发公司，而是一家软件测试公司。”足见其对于软件测试的重视程度。一般而言，为了保证复杂软件的正确性，从软件需求分析到软件方案设计，再到软件开发等不同生命周期，都会有测试环节。思考在软件不同生命周期中，软件测试所起到的不同作用。

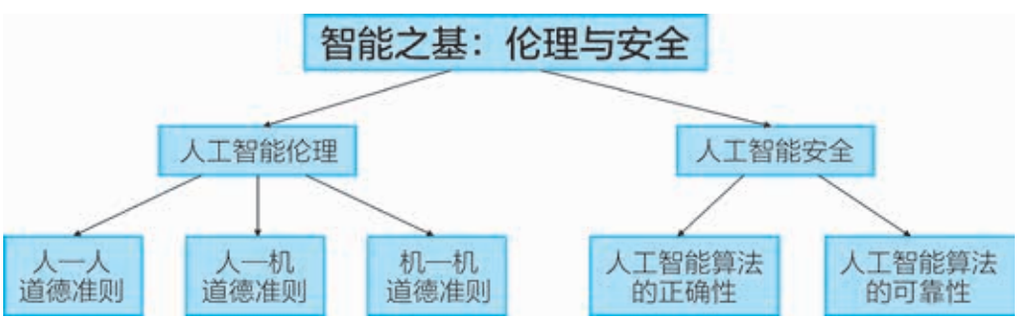
- “杀手机器人”指的是在没有人类决策和干预的情况下，机器人自主进行选择，完成目标寻找、定位和杀戮。许多人工智能和机器人研究人员宣布联合抵制“杀手机器人”的研制。思考并列举“杀手机器人”给人类社会带来的伦理道德冲击。

- 人们常说“魔高一尺，道高一丈”，人工智能安全防护技术随着攻击技术的不断进步而不断进步。必须意识到人工智能安全上的缺陷，才能最终解决人工智能存在的缺陷。辩证地思考人工智能安全防护技术在与攻击手段对决中如何不断提升。

学习目标

1. 了解人工智能伦理的基本概念和范畴。
2. 对涉及人工智能伦理的案例具有一定的辨析能力。
3. 掌握程序正确性证明的基本概念和保障信息系统安全的基本手段。
4. 明确人工智能社会化过程中应遵守的规范和法规。

内容总览



5.1 概 述

伴随着人类社会的发展进程，伦理学是介于美德和法律制度之间的规范群体行为的社会基本机能。美德是至圣的行为典范，法律制度惩罚罪恶。伦理规范常人行为，使之在实践中遵循社会普遍接受的准则。

随着人工智能的发展，机器与人类一道，承担着越来越多与人类社会息息相关的决策任务，由此引发了许多关于社会公平、伦理道德的新问题。

人工智能技术正变得越来越强大，那些最早开发和部署人工智能的企业现在也开始公开讨论其创造的智能机器给伦理道德带来的挑战。

国家《新一代人工智能发展规划》中明确指出：把握人工智能技术属性和社会属性高度融合的特征。既要加大人工智能研发和应用力度，最大程度发挥人工智能潜力，又要预判人工智能的挑战，协调产业政策、创新政策与社会政策，实现激励发展与合理规制的协调，最大限度防范风险。

人工智能伦理是一个正在兴起的研究领域，是一门研究人工智能理论的哲学思想基础、人工智能实践的价值观、机器伦理以及人工智能从业人员行为规范的实践伦理学科。传统伦理是研究人与人以及人与自然的关系和处理这些关系的规则，人工智能伦理是伴随着人工智能的理论与实践的出现而产生的，其主要研究对象不仅是人与人或人与自然，而且是人—机、机—机以及人—机混合系统之间应该遵守的准则，使得人机共融社会能够和谐发展。值得注意的是，这里的“机”可理解为算法或系统。人工智能伦理需要使人倡导的价值取向与伦理规范得以嵌入各种智能算法或系统中，使其遵守道德规范并具有一定程度的自主伦理抉择能力。为了实现人类的伦理价值，确保人工智能体的行为对人类透明、负责和可问责，需要适合于人工智能体的法律规范以及相应的技术手段。

如果说人工智能伦理是人工智能发展的“软规范”，那么保障人工智能安全的技术手段则是人工智能发展的“硬外壳”。人工智能面临的安全风险主要有两种类型：一是因人

拓展链接

伦理、工程伦理和人工智能伦理

伦理研究人与人和人与自然之间应该遵守的关系总则，工程伦理研究工程项目（如水电站建设、工业污染处理和基因克隆等）中应该维护的道德价值，人工智能伦理则是研究因为人工智能算法进展而形成人机混合系统或人机混合形态中人与人、人与机、机与机应该遵守的原则，如“自动驾驶车辆在要发生事故时，是优先保护路人还是车上的乘客”。我们需要明确，任何一项技术创新都要经历一个从不完善到完善、从不可控到可控的迭代过程。

工智能本身技术缺陷而潜在的内部隐患；二是对人工智能系统在实际环境中所要面临的外部攻击威胁。前者主要是因为人工智能技术本身不可避免地存在局限，应对实际情况时可能会出现一些“失误”，从而引发意外事故，因此需要保障人工智能系统的“正确性”；后者主要指网络黑客或破坏者可能对人工智能系统进行攻击，一旦攻击得逞，将造成严重的安全事故，因此需要保障人工智能系统的可靠性（即鲁棒性）。

5.2 人工智能伦理

按照科学哲学家波普尔的观点，世界由三部分组成，即第一部分的物理世界、第二部分的心理世界和第三部分的人工世界。人工智能时代正在构造由信息空间—物理空间—人类社会相互融合而形成的三元空间（Cyber-Physical-Society, 简称CPS），使得人—机—物紧密结合在一起，CPS的出现使得制定道德准则变得尤为重要。

1942年，阿西莫夫在科幻短篇小说《环舞》中首次提出了三条定律：（1）机器人不得伤害人类，或因不作为使人类受到伤害；（2）在不违背第一定律前提下，机器人必须服从人类的命令；（3）在不违背第一和第二定律的前提下，机器人必须保护自己。后来，阿西莫夫又加入了一条新定律：第零定律，即机器人不得伤害人类整体，或因不作为使人类整体受到伤害。

当前，随着人工智能迅猛发展，我们已经不能独立于机器人来制定道德准则，而是应该认真考虑人机共融社会形态的道德准则。人工智能伦理问题已经得到了世界各国的关切。

美国在2016年10月发布的《国家人工智能研究与发展战略计划》中确定了七个重点战略方向，其中就包括确保人工智能系统的安全性以及理解和应对人工智能带来的伦理、法律和社会影响。

中国在2017年7月向全社会公布的《新一代人工智能发展规划》中指出：人工智能是影响面广的颠覆性技术，其发展的不确定性将对就业、法律、伦理、隐私、国际关系等带来新挑战。《新一代人工智能发展规划》规划国家未来15年的人工智能发展，注重发展与规制平衡，强调把握人工智能技术属性和社会属性高度融合的特征，既要最大程度发挥人工智能潜力，又要加强前瞻预防与约束引导，确保走上安全、可靠、可控的发展轨道。

英国在2018年发布了《英国人工智能发展的计划、能力与志向》，提出了五条“人工智能准则”（AI Code）：发展人工智能，人工智能要遵循可理解性和公平性的原则，人工智能不应该被用来削弱个人、家庭或社区的数据权利或隐私权，所有公民都应在精神、情感和经济上与人工智能并驾齐驱，伤害、摧毁或欺骗人类的自主权力永远不应被赋予人工智能。

《麻省理工学院技术评论》在2018年3月发表了一篇文章《如果人工智能最终杀死了一个人，谁该负责》。这篇文章提出了一个严肃的问题：如果2023年自动驾驶汽车在全美广泛使用，其中一辆自动驾驶汽车撞击并杀死了一个人，应该采用什么样的法律呢？

下面介绍几个由人工智能所引发的案例，涉及人—机道德准则。

案例一：自动驾驶汽车逐渐进入社会实体

19世纪时，汽车刚被发明出来，它被认为是一种怪物，不少人认为汽车危险重重。因此，1865年英国议会通过了一部《机动车法案》，后被人嘲笑为“红旗法案”。这个法

案规定：每一辆在道路上行驶的机动车，必须由3个人驾驶，其中一人必须在车前面50m以外做引导，并且这个人还要用红旗不断摇动为机动车开道（提示人们危险将近），并且汽车速度不能超过6.4km/h。“红旗法案”放大了新生事物的危险性，在一定程度上阻碍了创新。

当前，自动驾驶汽车已被容许在部分城市的一些路段进行测试或行驶，开始进入人类社会。2018年4月，《智能网联汽车道路测试管理规范（试行）》颁发，首次从国家层面就规范自动驾驶道路测试做出规定。2018年6月，美国加州批准在获得载客运输许可和加州机动车辆管理局测试许可的前提下，自动驾驶汽车可载客运行。

但是，“自动驾驶”进入社会并非一帆风顺，还有一段漫长道路。2016年5月，约书亚·布朗在使用特斯拉 Model S 的自动驾驶功能时发生了车祸，并因此遇难，成为“自动驾驶”汽车进入社会发展中的重要事件。在历经漫长的调查后，这一事故最后归责为驾驶员过于信任人工智能系统，未能随时准备好“接管”自动驾驶系统可能出现的误判，结果“自动驾驶”系统得以免责。

2018年3月，一辆自动驾驶的Uber汽车在美国亚利桑那州撞死了一名女子。当时Uber 车辆处于自动驾驶模式，车上有一名安全员司机，遇难的女子当时推着自行车突然从阴影处走到车道上。调查结果显示：车子制导系统在撞到该女子6秒前发现了她，但是“为了降低车辆不稳定行为的可能性”，紧急制动策略没有起作用。

可以预见，在不久的将来，无人驾驶的人工智能算法会不断演进，厂商、使用者、算法、传感器等都是演进整体中的各个个体，于是并不存在对整体负责的“主体”，因此在伦理上可能面临根本的挑战：如何判断道德主体且让其负起应有的责任？

案例二：脸书数据泄露事件

英国资料分析机构剑桥分析（Cambridge Analytica）获取社交平台脸书（Facebook）大量个人资料，借以影响美国总统选举，引起社会关注和愤怒。如果掌握数据的公司或个人，利用手中的数据，刻意影响公众舆论、政治生态和社会安定，甚至影响军事战争决策，那就不亚于错误使用核武器了。如果互联网平台不加强管理，别有用心的人甚至可以在互联网平台上贩卖毒品和枪械以及散布恐怖信息，使得本来给大家进行信息交流的便利之器就变成了“罪恶之术”。

案例三：大数据“杀熟”

电商平台向不同熟客推荐价格更高的高端产品或服务，甚至给老顾客更高的报价，这叫大数据“杀熟”。

对大数据“杀熟”的关注，成了大数据人工智能带给人们便利过程中的一丝不愉快的杂音。消除社会公众对大数据被滥用的关注，避免“技术的贪欲”，是人工智能时代不可回避的现实选项。

人工智能通过对海量数据的梳理，让营销企业具备无限提升效率和精准服务的可能。但是，现在的网络平台，却借助大数据技术，对消费者精准靶向营销，不同用户不同定

价，特别是一些对价格不敏感的消费人群，溢价提供服务，从而出现了越是老用户价格越高的怪象。很多人在不知情的情况下，被“最懂我的人伤害”。

数字经济的问题，背后是技术伦理的准备不足；广泛共享的大数据，需要健康平衡的数字化生态来规范。

拓展链接

道德图灵测试

人工智能技术手段是解决“能还是不能”的问题，人工智能伦理则是在探讨“应该还是不应该”的问题。主流伦理学认为，一个合格的道德主体至少拥有意向性和一定程度的自由，而技术客体由于缺乏自由意志和意向性，因此无法对它们的行为负责。2000年，科林·艾伦、加里·瓦纳以及杰森·泽欣尔提出了道德图灵测试（Moral Turing Test，简称MTT），对道德自主体（Automated Moral Agent，简称AMA）的道德伦理进行测试。如何让具有高度自主性的机器人成为道德主体（道德机器），这是伦理学、哲学、认知科学、人工智能等研究面临的难题。

思考与练习

1. 为了让越来越具有自主能力的机器遵守人类道德伦理标准，我们需要对自主机器进行人工干预，或者将人类道德伦理准则转变成“代码”输入机器。请从伦理角度为无人驾驶机器人设计若干可代码化的道德准则。

2. 构造一个道德图灵测试的环境，针对某一类应用来测试机器的道德伦理水平。

5.3 人工智能安全

人工智能应用普及和发展的关键与安全息息相关。例如，执行关键性任务的人工智能应用、个性化推荐和城市大脑等，这些都需要人工智能系统具有很强的安全性。

人工智能安全包括两个方面：（1）人工智能算法的正确性，即其能够按照算法设计或算法目标来严格执行；（2）人工智能系统的可靠性，即其能够抵御外界的攻击，使得攻击者无法通过破坏和控制人工智能系统本身，或者通过特意改变输入来使系统不知不觉地做出攻击者想要的决定。如在高可靠性保障下，黑客在攻击无人驾驶系统时就无法造成“前方无障碍物”的错觉假象。

5.3.1 人工智能算法的正确性

一个人工智能算法所能体现的效果就是代码按序运行的结果。当编码好一个人工智能算法后，如何保证该算法能够顺畅运行、符合设计者预期来输出结果呢？这就涉及程序的正确性证明问题。

比如，编写了如下人工智能算法：（1）任意输入两个数，输出较大的数；（2）输入一幅图像，判断其是不是一幅人脸图像。

为了证明第一个算法的正确性，可以输入成千上万对数或者输入典型测试用例，检测该算法是否输出每一对数中较大的数。但是，即使所有测试用例均得到了正确结果，也无法保证程序是正确的，因为可能某个极端条件下的测试用例没用来检测算法，从而没有发现算法的“Bug”。

拓展链接

程序中的Bug

“Bug”一词指计算机或计算机程序出现的“故障”和“缺陷”。1947年9月9日，哈佛大学测试马克II型艾肯（Mark II Aiken）中继器计算机时，操作员在电路板中编号为70的中继器触点旁发现了一只飞蛾，它使得计算机无法正常工作。清除了这只飞蛾后，操作员把这只飞蛾贴在计算机日志上了，并写下了“首个发现bug的实际案例”字样，如图5.3.1所示。同时，发明了一个新词“debug”（调试）来表示计算机的故障排除。1968年软件危机的出现，进一步说明了对复杂程序中潜在故障进行排除（debug）的重要性。

图5.3.1　从Mark II Aiken计算机中去除的飞蛾及计算机日志（上面写着：First actual case of bug being found）

为了证明第二个算法的正确性，可以输入成千上万幅不同条件下所采集的人脸图像，如正常的人脸、戴黑色墨镜的人脸、强烈光线照射下的人脸等，对算法不断测试，然后加以改进。

上述正确性证明方法可称为程序测试（Programming Testing）。可以想象，为了找到程序的不足，要尽可能设计能暴露错误的测试用例，如边界条件用例等。

采用测试方法来验证程序的错误，只能保证在这些测试用例下，程序没有出错，却不能证明程序本身没有错误。因此，为了保证程序的正确性，必须从理论上研究“程序正确性证明”的方法，即通过形式化方法来证明程序的正确性。

早在20世纪40年代，图灵就提出了进行程序正确性证明和程序验证的想法。在这个想法中，图灵认为：假设在实现某功能的程序中，存在变量$x_1, x_2,\cdots, x_n$，输入谓词$P(x_1, x_2,\cdots, x_n)$与输出谓词$Q(x_1, x_2,\cdots, x_n)$。如果能证明该程序在运行谓词$P(x_1, x_2,\cdots, x_n)$之前是成立的，且在程序运行完毕后，谓词$Q(x_1, x_2,\cdots, x_n)$成立，那么程序正确性就被证明了。

以求两个数中较大数的程序为例，输入谓词为$P(x_1, x_2)$（x_1和x_2为任意两个数），输出谓词如下：

$Q(x_1, x_2) = x_1 \wedge x_1 > x_2$ or $Q(x_1, x_2) = x_2 \wedge x_2 > x_1$　　//输出x_1和x_2中较大数

输入的两个数相等时，发现输出谓词逻辑没有考虑 = 的情形，使得程序不正确。因此可以将谓词逻辑修改如下：

$Q(x_1, x_2) = x_1 \wedge x_1 > x_2$ or $Q(x_1, x_2) = x_2 \wedge x_2 > x_1$ or $Q(x_1, x_2) = x_1 \wedge x_1 = x_2$
//输出x_1和x_2中较大数，若两者相等，则输出其中一个值

从这里可以看出，为了对程序进行形式化证明，需要将代码序列转换成逻辑序列，然后加以证明。考虑到一个程序中有存在与运行状态相关的指令（如循环或选择等），很难将一个程序完全转换为逻辑序列，因此程序形式化证明目前仍然是一个尚待攻克的难点问题。

拓展链接

程序正确性证明

程序的正确性证明是计算机理论中的一个难点问题。图灵在1949年发表了一篇关于软件测试的文章，最早提出了通过断言（assertion）方法进行程序证明的思路，即可以通过证明程序里一系列断言的正确性来证明一个程序的正确性。随后，麦卡锡（1971年图灵奖获得者）、迪科斯彻（1972年图灵奖获得者）、罗伯特·弗洛伊德（1978年图灵奖获得者）、东尼·霍尔（1980年图灵奖获得者）、彼得·诺尔（2005年图灵奖获得者）等均对这个领域做出了杰出贡献。他们的基本思想是将程序算法本身转换成逻辑表达形式，进而应用形式化方法来证明程序本身是否正确。

5.3.2 人工智能算法的可靠性

在由信息空间—人类社会—物理空间组成的三元空间中，信息安全包括用户安全、内容安全、网络安全和系统安全，如图5.3.2所示。

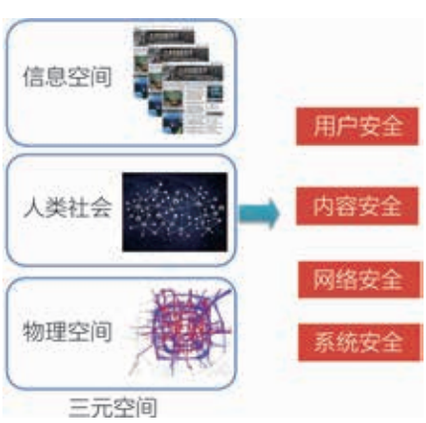


图5.3.2 信息安全示意图

用户安全指的是用户只能访问其被授予权限的系统，不能超越其权限来操作系统。内容安全指的是系统中代码和数据是正确的，未受到恶意修改。网络安全指的是信息在传输过程中未被截获泄密。系统安全是指保障整个信息系统从数据采集到数据分析以及决策行动等方面的安全性。

从前面章节的学习中，我们知道在通过标注大数据训练下，人工智能算法已经可以准确地识别图像中的视觉对象（如人脸、汽车等）和音频中的语音。

如果在正常样本中人为故意掺入噪声干扰以误导智能算法，使智能算法产生错误结果，这种在正常样本中添加了噪声干扰以误导智能算法的样本，被称为“对抗样本”（adversarial example）。

如果恶意使用对抗样本，可以欺骗自动驾驶汽车，使其不能正确识别道路停车标志，从而引发事故；可以欺骗语音识别系统，让系统误认为是“主人语音”或执行虚假命令；可以入侵城市交通系统或武器系统。这种通过对抗样本对智能系统发起的欺骗被称为“对抗性攻击”。

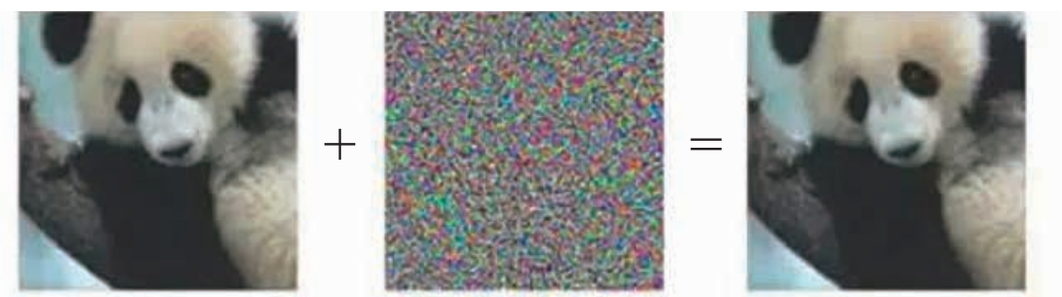


图5.3.3　对抗样本及其错误辨识结果

（左边：熊猫；中间：噪声图像；右边：加入噪声后的图像被智能算法辨识为长臂猿）

图5.3.3给出了一个对抗性攻击的例子。甲图是神经网络智能程序以57.5%置信度识别为熊猫的图像，乙图是噪声图像，丙图是通过甲、乙两图叠加得到的图像。人类可明了辨识如图5.3.3丙所示的是熊猫，但是智能程序将其识别为长臂猿，可能的原因是神经网络对图像识别途径与人脑对视觉信息认知不同所导致的。

实践与体验

对加入噪声后的图像进行分类

体验“对抗样本”对人工智能模型可靠性的影响。你可以自行搜集图片，使用ResNet，观察在加入噪声前后，分类结果的不同。

实践内容：

1. 收集图片数据，加入噪声。
2. 在Keras上构建和使用ResNet分类模型。
3. 收集实验结果，进行分析和解释。

实践步骤：

1. 收集图片数据，加入噪声。

你可以自行搜集数据，也可以使用3.2节“实践与体验”中的数据。加入噪声的方法如下：

（1）安装scikit-image。

scikit-image是一种方便操作图片数据的Python库。安装代码如下：

```
pip install scikit-image
```

（2）scikit-image 中给图片加入噪声的函数，代码如下：

```
skimage.util.random_noise(image, mode='gaussian', seed=None, clip=True, **kwargs)
```

2. 在 Keras 上构建和使用 ResNet 分类模型。

参考 3.2 节的“实践与体验”。

结果呈现：

多尝试几种不同类型的图片，并对每一张图片多尝试几次（噪声的产生是随机的），比较加入噪声前后图片分类结果的不同。

思考与练习

某程序员编写了如下人工智能程序：输入三个整数，输出其中所有的素数（质数）。为了测试该程序的正确性，准备了如下三个测试用例：（–3, 2, 2）、（π, 3, 5）和（5, 7, 19），程序分别输出：没有素数、输入错误、（5, 7, 19）。在这三个测试样例中，程序是正确的。

现在需要重新选用几个样例来测试程序的正确性，你会采用哪些特殊的样例？在这些样例测试下，程序输出均是正确的，这样就能确保程序的正确性吗？

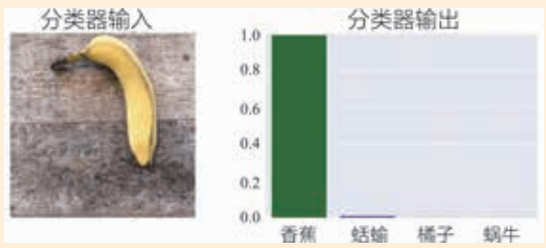
巩固与提高

1. 假设一个神经网络具有除法操作的功能，现在要对该神经网络这个功能的正确性进行验证，即输入两个数x和y，输出$\frac{x}{y}$的值。请设计一系列的测试用例来验证该神经网络执行功能的正确性。你设计的用例组合应当可以验证程序的以下方面：（1）对于正常的输入数据可以输出正确的结果；（2）可以发现输入格式中的错误并给出提示；（3）当除数为零时，可以发现错误并给出提示；（4）当除法答案数值过大无法在计算机中通用给定格式正确表示时，可以发现问题并给出提示。

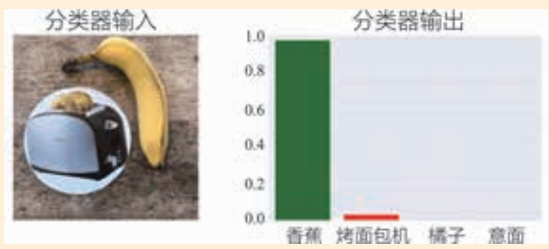
2. 2016年3月23日，微软在社交网络上推出了一个聊天机器人。微软希望该聊天机器人与网友进行对话，并用这些对话数据进一步训练机器人来提高它的交流能力。然而，由于相当多的用户刻意使用种族歧视和不礼貌的言语跟机器人聊天，使得机器人也开始“学会”使用这些言语聊天。这导致微软不得不暂时关停了该聊天机器人。请从人工智能安全方面（如输入数据对特定任务的可靠性）思考应该如何避免这种情形的发生。

3. 2017年12月，谷歌的研究团队提出了一种称为“对抗性数据子块”（Adversarial Patch）的技术，能够有效干扰进行图像识别的人工智能算法模型，使之出现误判。在这种技术中，只要将特定的对抗性数据子块放在待识别物体的旁边，就能够诱导人工智能算法模型忽略需要识别的物体而被对抗性数据子块吸引，做出错误的判断。对抗性数据子块只需占据原始待识别图像不到10%的面积，就能以90%以上的成功率使得模型做出误判。

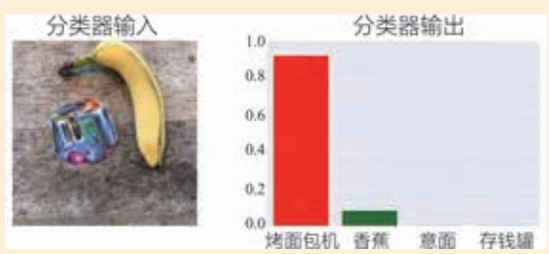
如图5.3.4所示，在香蕉旁边分别放置印有烤面包机的圆盘和对抗性数据子块时，根据图像分类模型的输出结果，可以看到对抗性数据子块显著干扰了图像识别模型的预测。



仅将香蕉作为图像分类模型的输入



将香蕉和烤面包机子图均作为分类模型的输入



将香蕉和烤面包机对抗数据子块均作为分类模型的输入

图5.3.4

试想一下，如果人工智能算法不能应对对抗性数据子块的攻击，那么当人工智能大规模应用于生活时，可能会产生哪些安全隐患？

项目挑战

人工智能伦理之辨：谁之过

人工智能推动了人类社会进步，出现了人机共融的社会形态。在这种社会形态中，机器代替人完成了一些原本由人完成的任务，如驾驶、生产和社会服务等。于是，用于人与人之间的传统道德准则发生了变化，需要认真考虑人机共融情况下道德准则出现的新问题。

比如说，两个人机共驾系统之间发生了伤亡事故，那么谁应该对这个事故负责呢？请对一个假想事例（应是在未来社会中可能会出现的）进行分析，给出评判理由。

项目任务

如图5.3.5所示，李某驾驶着某公司生产的配置了自动驾驶装置的汽车在A市道路上行驶。王某驾驶着一辆经交通管理部门核准的自动驾驶车辆在A市道路上测试。当天A市下雨，天气状况一般。

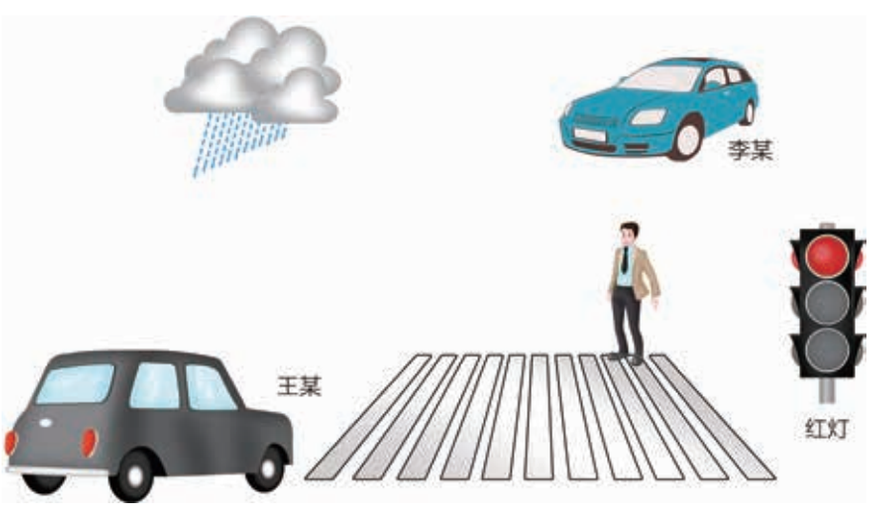


图5.3.5　交通出行示意图

在经过一个红绿灯时，为了避让一个闯红灯的行人，李某和王某车子相撞，不仅造成行人重伤，而且他们也受伤了。

请你根据上述描述，给出这次事故责任的认定报告，与其他分组进行辨析。

具体要求如下：

在这个事故中，涉及自动驾驶汽车的性能是否足够好（如在不同的天气情况下识别红绿灯和行人等）、不同公司的自动驾驶汽车是否有安全保障、驾驶无人车的驾驶员是否

及时对自动驾驶进行了干预、批准无人汽车上路测试是否合理以及行人闯红灯的状况等因素。请列举出各个方面应该承担的责任，并给出理由。

过程与建议

1. 分析已有案例

目前，已经出现了一些与无人驾驶汽车有关的交通事故，分析这些事故的成因、最终的事故认定结果，以及对已有事例判断的局限性。

2. 理解人机共驾系统中驾驶员、机器和行人的各自职责

在道路交通中，驾驶员、驾驶装置和行人都要对道路的安全负责，给出其应该承担的职责列表。比如：自动驾驶装置是否足够安全？李某和王某是否遵守了人机混合驾驶所规定的要求（如需要在危险状况下能够及时采取措施，对自动驾驶装置进行干预）？允许自动驾驶车上路测试是否合理？行人的行为是否构成全责？自动驾驶装置中算法供应商（如视觉识别系统）或控制系统供应商的产品是否足够安全？

3. 自动驾驶汽车性能失效分析

分析自动驾驶汽车在什么情况下，其性能会降低或失效，比如视觉识别系统在恶劣天气环境下性能会急剧下降。分析这些急剧下降的性能会对驾驶安全产生怎样的影响。

4. 辨析过程

对于在交流过程中不同的观点，需要为自己的观点提出支撑性的理由。

5. 撰写项目报告

基于以上工作，撰写分析报告，包括如下内容：

（1）事故过程分析。

（2）事故中所涉及各个方面应该承担的责任。

（3）对所涉及的责任主体未来发展方向的建议。

评价标准

请根据项目实施的过程、效果以及成果展示交流的结果，对自己完成项目的情况进行客观的评价，并思考后续完善的方向。把评价结果和完善方案填写在下页的表格中。

评价条目	说明	评分（1～10分）	评分主要依据阐述	后续完善方向
项目理解	在项目开展和分析报告中，对人工智能伦理理解正确、对案例任务的要求理解明确			
小组辨析	小组分工合理，协作紧密，合作有成效			
分析论辩结果的合理性	证据完整、论辩可信			
为智能社会中各类主体给出伦理道德建议	对人、机等智能主体在未来人机共融社会中发展的建议合理			
展示交流	展示经过精心准备，表达清晰、便于受众理解			