普通高中教科书

XINXI JISHU

信息技术

选择性必修 4

人工智能初步

教育科学出版社

普通高中教科书

XINXI JISHU

信息技术

选择性必修 4

人工智能初步

教育科学出版社
·北 京·

总 主 编 李 艺 董玉琦
本册主编 张剑平
本册副主编 余燕芳
主要编者 张剑平 余燕芳 林 斌 陈美锭 白晓东

出 版 人 李 东
责任编辑 贾立杰
版式设计 国美嘉誉文化 王 辉
责任校对 贾静芳
责任印制 叶小峰

普通高中教科书

信息技术 选择性必修 4 人工智能初步

教育科学出版社出版发行
(北京·朝阳区安慧北里安园甲 9 号)
邮编：100101
总编室电话：010-64981290 编辑部电话：010-64989637
出版部电话：010-64989487 市场部电话：010-64989009
传真：010-64891796
网址：http://www.esph.com.cn
各地新华书店经销
河南省四合印务有限公司印装
开本：890 毫米 ×1240 毫米 1/16 印张：9
2020 年 1 月第 1 版 2021 年 12 月第 5 次印刷
印数：1—500 册

ISBN 978-7-5191-1308-7
定价：15.50 元 (含光盘)
冀价审〔2022〕016039 价格举报电话：12315

图片来源：高品（北京）图像有限公司

前　言

近年来，随着互联网和大数据应用的快速发展，人工智能技术正在迅速走进人们的日常工作与生活。在巨大的潜在市场价值面前，全球互联网企业纷纷卷入其中，如谷歌、微软、苹果、百度、腾讯等都在积极进行人工智能领域的科学研究与产品开发。从谷歌的智能软件“阿尔法围棋（AlphaGo）”以4∶1战胜韩国围棋名将李世石，到腾讯的写稿机器人“梦幻写手（Dreamwriter）”在国家统计局公布年度居民消费价格指数（CPI）的第一时间就写出新闻稿，人工智能已经实实在在地来到了我们面前。

所谓智能，通常是指人类大脑的高级活动，人工智能则是针对人类智能行为的模拟、延伸和扩展的技术应用。如果说300多年前蒸汽机的出现解放了人类的体力，那么人工智能的应用则有望进一步解放人类的智力。

同学们一定想知道：什么是人工智能？它的原理是怎样的？什么是专家系统？机器学习是如何进行的？如何设计并开发一个简单的人工智能应用模块？人工智能的发展会面临哪些挑战？本书将带领大家寻找以上问题的答案，学习和体验人工智能应用，畅游人工智能的海洋。

本书包括“初识人工智能”“知识表示及其应用”“机器学习与深度学习”“体验人工智能应用”和“人工智能的辩证思考”五个单元。通过学习本书，同学们可以了解人工智能的基本概念与发展历程，知道知识表示、神经网络、机器学习等人工智能方法的原理与内涵，亲历设计与实现人工智能应用系统的基本过程，了解社会智能化所面临的伦理及安全挑战，辩证地思考人工智能的未来价值和潜在威胁，从此形成对人工智能的深刻理解。

希望同学们立足于本书，又不局限于本书，做到理论联系实际，在实践中练就技术本领，在学习中体验收获与分享，更学会在人工智能的发展中创新、创造！

目　录

第 1 单元　初识人工智能

2016年3月，AlphaGo（阿尔法围棋）对战人类棋手、世界围棋冠军、职业九段选手李世石，并以4∶1的总比分获胜；而在2017年10月，AlphaGo Zero（阿尔法零）又以100∶0的比分击败AlphaGo，迅速刷新人工智能水平的高度，使人类棋手无法企及。在信息时代的今天，人工智能既不断给我们惊喜又无处不在，每天伴随着我们的工作、学习与生活。人们可以用指纹或虹膜来打卡签到；可以用OCR软件来识别文档；可以用语音助手来订飞机票、订火车票、订餐或订酒店；可以用拍照搜索功能在网上查找需购买的物品、搜索要学习的课程等。

了解人工智能的基本概念、基本特征及其发展历程，是学习人工智能的第一步。在本单元中，我们将围绕“了解人工智能的基本概念、基本特征及起源与发展”项目开展学习，除通过数字化学习了解人工智能的基本概念及发展历程外，还将通过体验人工智能典型应用的实践活动来认识人工智能的基本特征。

为了完成该项目，需要思考以下问题：什么是人工智能？它的基本特征是什么？它的发展历程是怎样的？人工智能有哪些典型应用？人工智能发展的主要驱动因素有哪些？为此，我们需要完成以下任务：

- ◆ 通过体验图灵测试及人工智能的典型应用了解人工智能
- ◆ 借助网络调查人工智能的起源、发展历程及当前影响人工智能发展的主要因素

1.1　人工智能的概念

当前的生活中，可以看到人们用语音识别来控制家电，用指纹识别来打开房门，用车牌识别来进入车库，用人脸识别来进行网银转账……，种种现象说明，语音识别、图像识别等人工智能（Artificial Intelligence，AI）技术的迅速发展，扩大了计算机技术的应用范围，提高了计算机与人的交互效率，为人们的工作与生活提供了更多的便利。可以说，人工智能时代已经到来了。

学习目标

★ 了解人工智能的基本概念与基本特征。

★ 体验图灵测试。

★ 体验人工智能的典型应用。

近来，人工智能这一名词经常出现在新闻报道中，大家对此十分感兴趣。那么，什么是人工智能？人工智能有哪些基本特征？让我们来了解一下吧。

任务　通过体验图灵测试及人工智能的典型应用了解人工智能

※ 活动 1　体验图灵测试

在开展活动之前，先要了解什么是图灵测试。图灵测试（The Turing Test）由英国科学家艾伦·麦席森·图灵发明，是一种在将机器与人相比较的思路下测试机器“智力”的方法。即将一台机器和一个人同时作为测试对象，使其与测试者们（人）相隔离：测试者们只能通过装置（如键盘）向测试对象提问题并根据所得回答判断哪个是真人、哪个是机器。进行多次提问及判断后，如果有超过30%的测试者不能准确分辨

出测试对象是人还是机器，则认为这台机器通过了图灵测试，认为它“有”了类似人类的智能。

本活动由“通过互联网与智能聊天软件进行交谈”和“通过图灵测试判断智能聊天软件是否具有智能”两部分组成。通过该活动，可以了解人工智能中的自然语言理解技术和图灵测试的方法。

◆ 通过互联网与智能聊天软件进行交谈

选择并安装一款常见的智能聊天软件（例如：微信或QQ中的聊天机器人软件），或者直接打开计算机设备中的智能聊天软件（例如：PC机Windows 10环境下的Cortana、iPhone手机上的Siri等），通过语音与计算机聊天，看看计算机是否能够理解人类的话语。计算机的反应与人类有何不同？计算机是否足够“聪明”？

图1.1.1所示是人类与智能聊天软件Cortana进行交谈的一个例子。

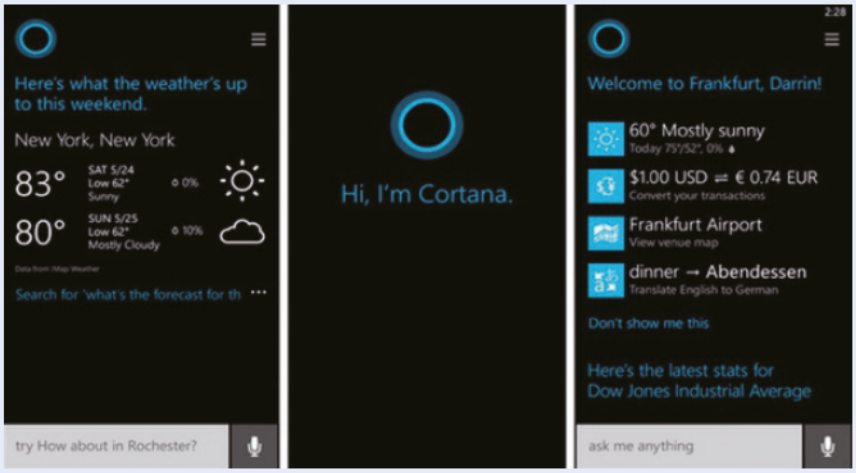


图1.1.1 与智能聊天软件Cortana交谈的例子

◆ 通过图灵测试判断智能聊天软件是否具有智能

借助一款智能聊天软件，师生一起玩游戏：学生甲和学生乙分别隔离在其他同学不能看到的空间里，回答由老师协助其他同学设计的一系列问题。学生甲和学生乙中的一人直接将自己的答案写出来通过投影展示给全体同学，另一人则扮演信使的角色，即先将问题输入智能聊天软件，待软件给出答案后，再转达给全体同学。一组对话之后，请同学们判断哪组答案来自真正的人、哪组答案来自聊天软件，并将判断结果通过“图灵测试管理系统”提交，提交后系统自动给出统计结果。教科书配套资源中提供的“图灵测试管理系统”界面如图1.1.2所示。

图1.1.2 利用“图灵测试管理系统”判断智能聊天软件是否具有智能

自然语言理解与智能聊天软件

自然语言理解（Natural Language Understanding）是一项人工智能技术。它研究如何使计算机理解和运用人类的自然语言，实现人机之间基于自然语言的交流，以代替人的部分脑力劳动，如查询资料、解答问题、摘录文献和汇编资料等工作。

智能聊天软件也称为聊天机器人，这是一类用来模拟人类对话或聊天的计算机软件，也是自然语言理解技术的主要应用实例之一。世界上最早的聊天机器人诞生于20世纪60年代，名为Eliza。

智能与人类智能

什么是智能（Intelligence）？关于这个问题，至今没有一个统一的答案。智能与智力相关。智力是运用知识解决问题的能力。智能是指知识的集合与能力的综合或总和，是静态的知识和动态的智力综合所体现出的一种能力。智能具有感觉、收集、汇集、理解和选择信息的功能。智能过程包括了感觉、记忆、思维、语言、自适应和行为的整个过程。

广义的智能包括人类智能、人工智能和集成智能。狭义的智能仅指人类智能，又叫自然智能，是指人在认识和改造客观世界的活动中，由思维过程和脑力活动所体现出来的智慧与能力。

人类智能主要包含三个方面：思维能力、感知能力和行为能力。在科学技术领域，智能也指机器所具有的自动控制能力和根据环境自我调节的能力。

集成智能是指基于人类智能和人工智能相结合的人—机系统所具有的智能。

什么是人工智能

人工智能兴起于20世纪50年代后期，作为一个不断发展中的领域，至今尚无统一的定义。斯坦福大学人工智能研究中心的尼尔逊教授给人工智能下了这样一个定义：“人工智能是关于知识的科学——怎样表示知识以及怎样获得知识并使用知识的科学。”而麻省理工学院的温斯顿教授认为：“人工智能就是研究如何使计算机去做过去只有人才能做的智能工作。”这些说法反映了人工智能的基本思想和所包含的基本内容。通俗地说，可以认为人工智能是研究如何让计算机去完成以往需要人的智力才能胜任的工作，也就是研究如何应用计算机的软硬件来模拟人类某些智能行为的基本理论、方法和技术。

人工智能是计算机科学的一个分支，它被誉为世界三大尖端技术（空

间技术、能源技术、人工智能）之一。与一般的信息处理技术相比，人工智能技术在求解策略和处理手段上都有独特的风格。目前，人工智能的不少研究领域如自然语言理解、模式识别、机器学习、数据挖掘、智能检索、机器人技术等都走在了信息技术发展的前沿，有许多应用成果已经进入人们的生活、学习和工作中，对社会发展产生了重要影响。人工智能的研究、应用和发展在一定程度上决定着信息技术发展的方向，信息技术的广泛应用也对人工智能技术的继续发展提出了强烈的要求。

人工智能的发展道路虽然起伏曲折，但是硕果累累。无论是基础理论创新，还是关键技术突破，抑或是规模产业应用，都精彩纷呈。几乎信息技术领域中的所有主题和热点，如智能硬件、无人机、机器人等，其发展突破的关键环节都与人工智能有关。

※ 活动2　体验人工智能的典型应用

按表1.1.1所示的主题划分小组，登录人工智能开放平台（如腾讯AI开放平台），在老师的指导与帮助下，体验人工智能的典型应用。

要求每个小组针对所选的主题开展上机实践与协作学习，并完成相关主题的知识获取、筛选与整理，然后以小组为单位在班级中进行分享与讨论。

上述活动完成后，请填写表1.1.1。

表1.1.1　体验人工智能的典型应用活动记录表

序号	主题	结果记录
1	人脸检测与分析	拍摄或上传一张个人正面照片，识别结果如下。 性别：男；年龄：16；表情：心花怒放；魅力：90
2	美食图片识别	
3	机器翻译	
4	语音识别	
5	其他	

人工智能开放平台

目前常见的人工智能开放平台有百度AI开放平台、腾讯AI 开放平台、阿里人工智能、网易人工智能、亚马逊人工智能服务、英特尔人工智能服务等。

腾讯AI 开放平台提供了OCR、人脸识别、图片特效、图片识别、敏感信息审核、智能闲聊、机器翻译、基础文本分析、语义解析、语音识别和语音合成等开放的人工智能典型应用服务。其中，手写体OCR的实例如图1.1.3所示。

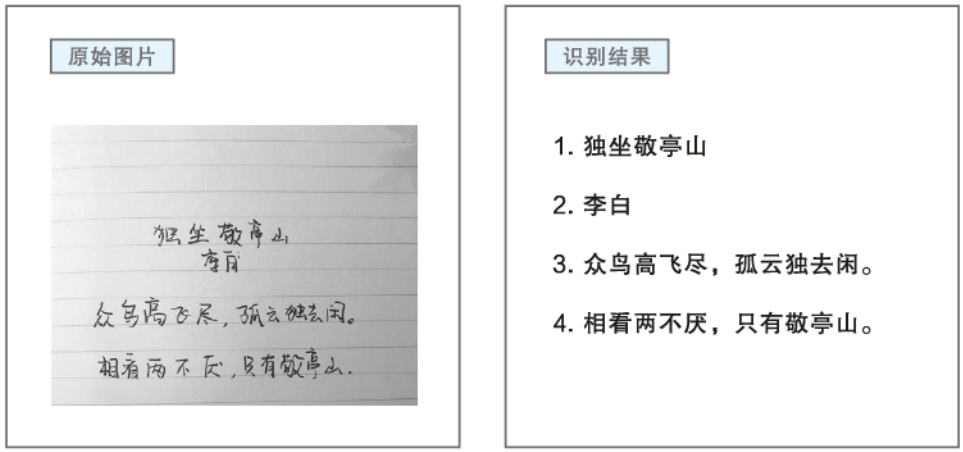


图 1.1.3　腾讯AI 开放平台的手写体OCR实例

身边的人工智能应用

语音助手

语音助手是一种智能型的手机应用，通过即时问答实现交互，帮助用户解决日常生活中的一些问题。语音助手能“听见”用户所说的话，进行语义判断并迅速做出回应，和用户进行语音聊天，或者听指令帮用户操控智能设备。语音助手被唤醒、听明白、会说话的过程，背后对应的是语音识别、语音合成等一系列智能技术。

智能手机中的语音助手实例如图1.1.4所示。

图 1.1.4　智能手机中的语音助手实例

扫地机器人

扫地机器人通常采用刷扫和抽吸相结合的方式，将地面杂物吸进内置的垃圾收纳盒，从而完成地面清理的任务。扫地机器人一般为圆盘形，可以通过遥控器或机器上的操作面板进行操作，也能够定时或预约时间打扫。扫地机器人以可充电电池为动力源，当电量不足时，能够自动寻找并行驶至充电口进行充电。扫地机器人的前方设有感应器，可侦测障碍物，碰到墙壁或其他障碍物时会自行避开，清扫路线的选择策略因各厂商的不同设定而有所不同。

近年来，扫地机器人已逐渐为人们所熟悉，成为家用电器中的新宠。典型的家用扫地机器人如图1.1.5所示。

图1.1.5　某品牌扫地机器人

智能导航

智能导航系统是利用卫星定位系统提供的位置、速度及时间等信息，配合高精度电子导航地图的路线规划能力，为用户提供智能导航功能的软件系统。它可以帮助用户准确、实时地在电子地图上规划行车路线，引导用户按规划的路线行驶。当位于某地的用户通过语音或其他输入方式向系统发出要到达的目的地指令后，该系统会迅速识别出用户的意图并规划出合理的行驶路线，在显示屏上实时显示车辆所在的位置、速度、限速、附近道口及周围设施等信息，配合语音向用户提示正确的行驶路线和方向，直至引导车辆到达目的地。这对驾车一族来说非常实用，对经常外出去陌生目的地旅游、探险、进行跨区商务活动的驾车者来说更是不可或缺。某智能导航系统的工作界面如图1.1.6所示。

图 1.1.6　某智能导航系统的工作界面

人脸识别

人脸识别是基于人的脸部特征信息进行身份识别的一种生物识别技术，也被称作人像识别或面部识别。通常用摄像机或摄像头采集含有人脸的图像或视频流，再对采集到的人脸信息进行处理实现身份识别。现今人脸识别技术已广泛应用于金融、司法、军队、公安、边检、政府、航天、电力、工厂、教育、医疗等众多领域。如人脸识别可以应用到电子政务和电子商务方面，在人脸识别技术成熟以前，审批或者交易的授权只能靠密码来实现，如果密码被盗，就无法保证信息安全，而使用人脸这种生物特征，就可以做到当事人数字身份和真实身份的统一，提高电子商务和电子政务系统的安全性与可靠性。人脸识别系统的人机交互界面如图1.1.7所示。

图 1.1.7　刷脸认证

拓展练习

你知道电影《星球大战》中的C-3PO、电影《终结者》中的T-850、电影《机械战警》中的墨菲吗？它们都是科幻电影中著名的智能机器人，承担了很多人类不易或者不能完成的任务。如今，智能机器人已经开始走出科幻作品，走进人们的生活。如果仔细观察，你会发现身边很多工作已经被人工智能设备所取代。走进医院，来到车站，你会发现许多原来是售票员、挂号员、收银员的位置已经被机器取代了，那么以后还会有哪些工作会被机器取代呢？问诊的医生会吗？咨询的律师会吗？又有哪些工作不会被人工智能取代呢？

在老师的指导下，通过数字化学习或走访医院、车站或银行等服务单位，了解今后很有可能被人工智能取代的工作，以及不大可能被人工智能取代的工作，填写表1.1.2，并在班级中进行分享与讨论。

表1.1.2 学习活动记录表

很有可能 被人工智能取代的工作	不大可能 被人工智能取代的工作	简述理由
智能超市的收银员与安保员	智能超市的软硬件维护人员	智能超市的收银与安保工作很容易被人工智能所取代，而智能系统本身的软件和硬件维护还是需要人工进行

1.2　人工智能的起源和发展

凡是过往，皆为序章。过去已经预先决定了后续将要发生的事。因此，研究过去是必要且重要的。人工智能如何发展到今时今日的过程和经验，将决定这个领域的未来。在人工智能领域，与未来相比，也许现在连真正起步都还算不上，我们还需要继续克服难题，不断深化研究。

学习目标

★ 了解人工智能的起源和发展。

★ 了解人工智能发展的主要驱动因素。

人工智能是什么时候开始出现的？时至今日，人工智能的发展历程中发生了哪些振奋人心的大事？让我们一起来回顾一下人工智能的起源和发展历程吧。

任务　借助网络调查人工智能的起源、发展历程及当前影响人工智能发展的主要因素

※ 活动 1　梳理人工智能发展的重大历史事件

按表1.2.1所示划分小组，在老师的指导下，设计出高效的搜索关键词，如达特茅斯会议；在老师的推荐下，选择权威性较高的官方网站，开展数字化学习，寻找人工智能发展中的重大历史事件。

要求每个小组展开协作学习，并完成相关主题的知识获取、筛选与整理，然后以小组为单位在班级中进行分享与讨论。

上述活动完成后，请填写表1.2.1。

表1.2.1 人工智能发展的重大历史事件

时间	重要事件
1950年	图灵提出了著名的图灵测试，图灵测试至今仍然是人工智能的重要测试手段之一
1951年	马文·明斯基（Marvin Minsky）建造了第一台神经网络机，并将其命名为SNARC（Stochastic Neural Analog Reinforcement Calculator）
1956年	达特茅斯会议是人工智能诞生的标志性事件，会上正式确立了人工智能的概念
1957年	
1969年	
1980年	
1994年	
1997年	
2006年	
2011年	
……	

人工智能的发展历程

从诞生至今，人工智能已有60多年的历史，大致经历了三次发展浪潮。第一次浪潮为20世纪50年代末至20世纪80年代初；第二次浪潮为20世纪80年代初至20世纪末；第三次浪潮为21世纪初至今。在人工智能发展的前两次浪潮当中，由于技术上未能实现突破性进展，相关应用始终难以符合人们的期待，无法支撑起大规模商业化应用，最终在经历过两次高潮与低谷之后，人工智能归于沉寂。随着信息技术的快速发展和互联网的快速普及，以2006年深度学习模型的提出为标志，人工智能迎来了第三次浪潮，人工智能得以高速成长，且已经产生了巨大的商业效益。人工智能的发展历程如图1.2.1所示。

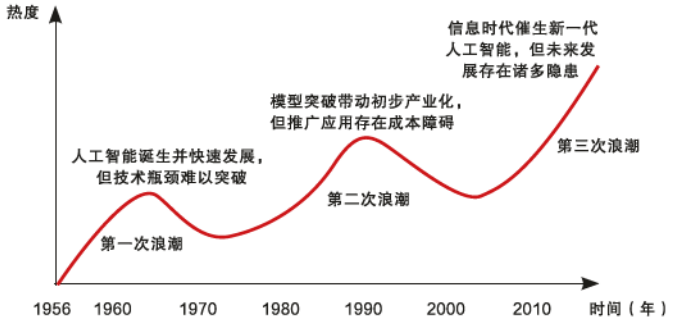


图1.2.1 人工智能发展历程示意图

第一次浪潮：人工智能诞生并快速发展，但技术瓶颈难以突破

1956年到1974年是人工智能发展的第一个黄金时期。科学家将符号方法引入统计方法中进行语义处理，出现了基于知识的方法，人机交互开始成为可能。科学家发明了多种具有重大影响的算法，还制作出具有初步智能的机器。人工智能发展迅猛，以至于研究者普遍认为人工智能代替人类只是时间问题。

1974年到1980年，人工智能发展的瓶颈逐渐显现，逻辑证明器、感知器、增强学习只能完成指定的工作，对于超出范围的任务则无法应对，智能水平较低，局限性较大。先天的缺陷是人工智能在早期发展过程中遇到的瓶颈，研发机构对人工智能的热情逐渐冷却，来自各方的对人工智能的资助也被缩减或取消，人工智能第一次步入低谷。

第二次浪潮：模型突破带动初步产业化，但推广应用存在成本障碍

进入20世纪80年代，人工智能再次回到了公众的视野当中。与人工智能相关的数学模型研究取得了一系列重大成果，卡耐基·梅隆大学为DEC公司制造出了专家系统，这个专家系统可帮助DEC公司每年节约4000万美元左右的费用，特别是在决策方面能提供有价值的支持。

为推动人工智能的发展，研究者设计了LISP语言，并针对该语言研制了LISP计算机。该机型计算机指令执行效率比通用型计算机更高，但价格昂贵且难以维护，始终难以大范围推广普及，逐渐被市场淘汰，人工智能硬件市场出现明显萎缩，专家系统也逐渐淡出人们的视野。同时，政府和社会各方投入经费的热情开始下降，人工智能又一次进入低谷。

第三次浪潮：信息时代催生新一代人工智能，但未来发展存在诸多隐患

随着互联网的普及、传感器的泛在、大数据的涌现、电子商务的发展、信息社区的兴起，数据和知识在人类社会、物理空间和信息空间交叉融合、相互作用，人工智能发展所处信息环境和数据基础发生了巨大的变化，这些变化构成了驱动人工智能走向新阶段的外在动力。与此同时，人工智能的目标和理念出现重要调整，科学基础和实现载体取得新的突破，类脑计算、深度学习、强化学习等一系列技术的萌芽也预示着内在动力的成长，人工智能的发展进入一个新的阶段。

人工智能水平快速提升，但有许多潜在的隐患。得益于数据量的快速增长、计算能力的大幅提升以及机器学习算法的持续优化，新一代人工智能在某些给定任务中已经展现出达到或超越人类的工作能力，如

AlphaGo就是在这个时候脱颖而出的；人工智能亦逐渐从专用型智能向通用型智能过渡，且有望发展为抽象型智能。随着应用范围的不断拓展，人工智能与人类生产生活的联系愈发紧密，一方面给人们带来诸多便利，另一方面也产生了一些潜在的问题：一是人工智能将取代很多工作岗位，结构性失业可能更为严重；二是隐私保护成为难点，数据拥有权、隐私权、许可权等界定存在困难；等等。

※ 活动2　调查当前推动人工智能发展的几个主要驱动因素

互联网催生了大数据，而大数据和计算机硬件的发展进一步推动了人工智能的快速发展。当前人工智能主要致力于训练机器（计算机）看懂图像、听懂语言、处理大数据，实现人机自然交互，辅助人类进行决策，最终实现自我决策。而支撑这一目的实现的主要驱动因素可以归纳为三点，即大数据、算法和算力。

按表1.2.2所示的主题划分小组，在老师的指导下，设计出高效的搜索关键词，如人工智能、大数据；在老师的推荐下，选择权威性较高的电子书籍，如吴军的《智能时代——大数据与智能革命重新定义未来》，开展数字化学习，了解大数据、算法、算力和人工智能的关系的相关知识。

要求每个小组对上述主题展开协作学习，并完成相关主题的知识获取、筛选与整理，然后以小组为单位在班级中进行分享与讨论。

上述活动完成后，请填写表1.2.2。

表1.2.2　大数据、算法、算力和人工智能关系的活动记录表

序号	主题	结果记录
1	什么是大数据	大数据（Big Data）是以容量大、类型多、存取速度快、应用价值高为主要特征的数据集合，它正快速发展为对数量巨大、来源分散、格式多样的数据进行采集、存储和关联分析，从中发现新知识、创造新价值、提升新能力的新一代信息技术和服务业态
2	大数据的来源	1. 机器上配备的监测装置产生的传感数据； 2. 互联网和其他使用者行动和行为的信息； 3. 使用者自身产生的数据和信息； 4. 音频、视频和符号数据； 5. 专业研究机构产生的大量数据
3	AlphaGo的算力	AlphaGo的早期版本（Fan）有40个搜索线程，1202个CPU（中央处理器），176个GPU（图形处理器）
4	AlphaGo的几种算法	
5	AlphaGo与棋谱数据的关系	

数据是指对客观事件进行记录并可以鉴别的符号，是对客观事物的性质、状态以及相互关系等进行记载的物理符号或这些物理符号的组合。它是可识别的、抽象的符号。知识是人类在实践中认识客观世界（包括人类自身）的成果，它包括事实、信息的描述或在教育和实践中获得的技能。它可以是关于理论的，也可以是关于实践的。

人工智能的发展几起几落，虽很大程度上受到业界对人工智能的认知变化的影响，但究其根本，还是在于技术与环境条件的发展和局限。没有配套的基础设施，有关人工智能的所有设想都将沦为幻想。人工智能发展的主要驱动因素如图1.2.2所示。

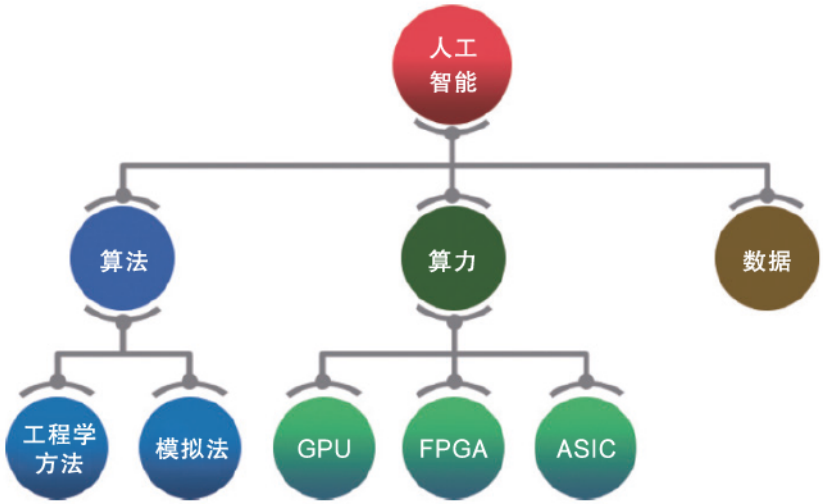


图1.2.2　人工智能发展的主要驱动因素

典型算法

遗传算法（Genetic Algorithm，GA）：是一类借鉴生物的进化规律演化而来的随机化搜索方法。它由美国的约翰·霍兰德（John Holland）于1975年提出，其主要特点是直接对结构对象进行操作，不存在求导和函数连续性的限定；具有内在的隐并行性和更好的全局寻优能力；采用概率化的寻优方法，能自动获取和指导优化的搜索空间，自适应地调整搜索方向，不需要确定的规则。遗传算法的这些性质，已被人们广泛地应用于组合优化、机器学习、信号处理、自适应控制和人工生命等领域。它是现代智能计算领域中的关键技术。

粒子群优化算法（Particle Swarm Optimization，PSO）：又称为粒子群算法、微粒群算法或微粒群优化算法，是通过模拟鸟群觅食行为而发展起来的一种基于群体协作的随机搜索算法。通常认为它是群集智能（Swarm Intelligence，SI）的一种。它可以被纳入多主体优化系统（Multiagent Optimization System，MAOS）。粒子群优化算法是埃伯哈特（R.C.Eberhart）和肯尼迪（J.Kennedy）发明的。

蚁群算法（Ant Colony Optimization，ACO）：是一种用来寻找优化路径的概率型算法。它由马克·多里戈（Marco Dorigo）于1992年在他的博士论文中提出，其灵感源于蚂蚁在寻找食物过程中发现路径的行为。这种算法具有分布计算、信息正反馈和启发式搜索的特征，本质上是进

化算法中的一种启发式全局优化算法。

算力举例

“谷歌大脑”用了上万个通用处理器“跑”了数天来学习如何识别猫脸；战胜韩国棋手李世石的AlphaGo则依赖于体积巨大的云服务器，这些服务器使用了上千个CPU和数百个GPU，在比赛时平均每局电费近3000美元。对于绝大多数智能需求来说，基于通用处理器的传统计算机成本高、功耗高、体积大、速度慢，令人难以接受。

目前，对于迅猛发展的人工智能来说，上层的应用都依赖于底层的核心能力，而这个核心能力就是人工智能处理器（芯片）的计算能力。如果在计算所需芯片技术上不能突破，人工智能应用就不可能真正成功。

GPU：图形处理器（Graphics Processing Unit，GPU），又称显示核心、视觉处理器、显示芯片，是一种专门在个人计算机、工作站、游戏机和一些移动设备上承担图像运算工作的微处理器。GPU的计算能力非常强，被广泛应用到图形图像处理、数值模拟、机器学习算法训练等任务中。在“谷歌大脑”的应用中，12个GPU可达到相当于2000个CPU的运算性能。

ASIC：一种为专门目的而设计的集成电路（Application Specific Integrated Circuit，ASIC），是指应特定用户要求和特定电子系统的需要而设计、制造的集成电路。ASIC的特点是面向特定用户的需求，与通用集成电路相比具有体积更小、功耗更低、可靠性与性能更高、保密性更强等优点，是一种比GPU更高效的芯片，但是其定制化也决定了它的可迁移性低，一旦专用于一个设计好的系统中，是不可能迁移到其他系统中的，并且其造价高，生产周期长，这些因素使得它在目前的研究中缺乏竞争力。

FPGA：现场可编程逻辑门阵列（Field－Programmable Gate Array，FPGA），是指可以通过软件手段更改、配置器件内部连接结构和逻辑单元，完成既定设计功能的数字集成电路。FPGA 最大的优点是动态可重配、性能功耗比高，非常适合在云端数据中心部署。全球七大超级云计算数据中心包括 IBM、Facebook、微软、AWS 以及 BAT（百度、阿里巴巴、腾讯）都采用了 FPGA 服务器。FPGA在GPU和ASIC间实现了平衡，很好地兼顾了处理速度和控制能力。

数据的产生与积累

在人工智能发展的早期阶段，获得大量的数据并非易事，甚至成为人工智能发展历程中致命的制约因素。但这一问题正随着大数据时代的到来而得到解决，特别是联网设备的爆发和服务生态的完善，使得数据

来源增多，促使数据呈指数式增长。

互联网的演进和催生的新业态，又进一步吸引了大量用户的积极参与，实现了数据“产生—使用—新数据产生—再使用”的闭环，这个闭环恰恰是人工智能自主学习和知识管理的基础。

● 结合AlphaGo Zero实例的讨论

AlphaGo Zero采用新的机器学习形式，它从零开始，面对的只是一张空白棋盘和游戏规则，从中自学围棋对弈的一招一式，在几天的时间里就可以极大地提高围棋水平。在一次对决中，AlphaGo Zero以100：0的不败战绩绝杀“前辈”——击败李世石的AlphaGo。

Zero的英文意思是“零”。除了最基本的围棋规则（棋盘的几何学定义、轮流落子规则、终局输赢计算、打劫等），AlphaGo Zero就是一张白纸。它不需要参考任何人类棋谱，完全自我学习，并且学习时间很短。它依赖于TPU的高速运行，几天时间就可完成数百万盘自我对弈学习，而一名职业棋手穷其一生，可能都没有完成1000局正式比赛。

原来的AlphaGo需要与人类专家进行成千上万次对弈，才能从中获取数据，AlphaGo Zero则截然不同。它从零开始，面对的只是一个空白棋盘和游戏规则。它无师自通，仅仅通过自学使自己的游戏技能得以提高。

AlphaGo Zero的无师自通主要表现为它可以根据规则自动产生数据。它首先随机下棋，然后根据输赢的准则来调整神经网络里的参数，通过不断的训练，接下来就不是随机下棋了，所有过程中产生的棋局数据，都被它拿去用来接受训练了。它是通过自我博弈而自己产生数据，用产生的数据来训练神经网络，从而提升技能。

AlphaGo Zero所采用的神经网络是一种新的强化学习算法，即自我对抗的竞争性训练。它不需要人类的样例或指导，不提供基本规则以外的任何领域知识，使用强化学习就能够达到超越人类的水平。因此，AlphaGo Zero需要一些基本的规则，表明不同状态下的下棋规则和输赢的规则。规则是一种知识形式，是高度提炼后的信息。AlphaGo Zero在训练过程中通过事先指定的规则进行不断的自我对弈，生成数据，优化搜索，从而完成自我进化。

从数据的角度看，AlphaGo Zero是根据知识产生数据，大数据处理则是根据数据产生知识。显然，两者具有很强的互补关系，如图1.2.3所示。

AlphaGo Zero这种新程序代表着人类在建造真正具有智能的机器方面向前迈了一步，因为即使在没有大量训练数据的情况下，机器也可以找出解决问题的方法。

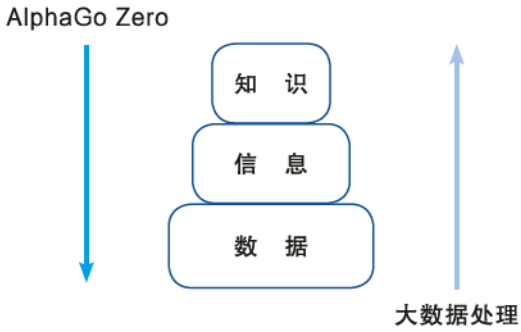


图1.2.3 AlphaGo Zero和大数据处理的关系

拓展练习

AlphaGo以4：1的总比分战胜人类职业棋手李世石。那么AlphaGo是怎么做到的？面对棋局时，AlphaGo通过评估和估值，从众多的可下子点中选择若干个“认为”最好的可下子点，同时判断棋局是否有利，从而从众多的搜索和模拟中计算出最优走法。

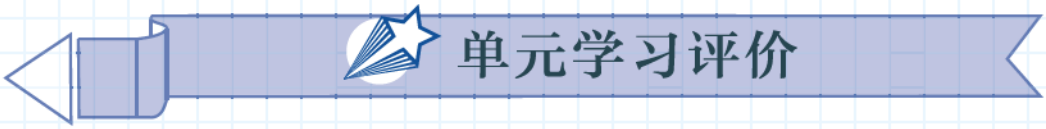
请按照表1.2.3所示划分小组，在老师的指导下，设计出高效的搜索关键词，如了解AlphaGo Zero；在老师的推荐下，选择较权威的官方网站，开展数字化学习，了解AlphaGo Zero的主要技术实现思路。

要求每个小组展开协作学习，并完成相关主题的知识获取、筛选与整理，然后以小组为单位在班级中进行分享与讨论。

上述活动完成后，请填写表1.2.3。

表1.2.3 AlphaGo Zero的主要技术实现活动记录表

序号	组成部分	功能与特点
1	走棋网络（Policy Network）	给定当前局面，预测/采样下一步的走棋
2	快速走子（Fast Rollout）	其目标和走棋网络的目标一样，但在适当牺牲走棋质量的条件下，速度要比走棋网络快1000倍
3	估值网络（Value Network）	
4	蒙特卡罗树搜索（Monte Carlo Tree Search，MCTS）	



单元学习评价

通过本单元的学习，我们了解了人工智能的概念，认识了人工智能的起源和发展，体验了图灵测试和人工智能的典型应用。你是如何看待人工智能的？你是否了解人工智能发展的重大历史事件，以及当前推动人工智能发展的几个主要驱动因素？请分小组交流并反思，开展自评或小组评价。

1. 图灵测试是由_____国科学家_____发明的，是一种在将机器与人相比较的思路下测试机器_____的方法。

2. 人类智能主要包含_____、_____和_____三个方面。

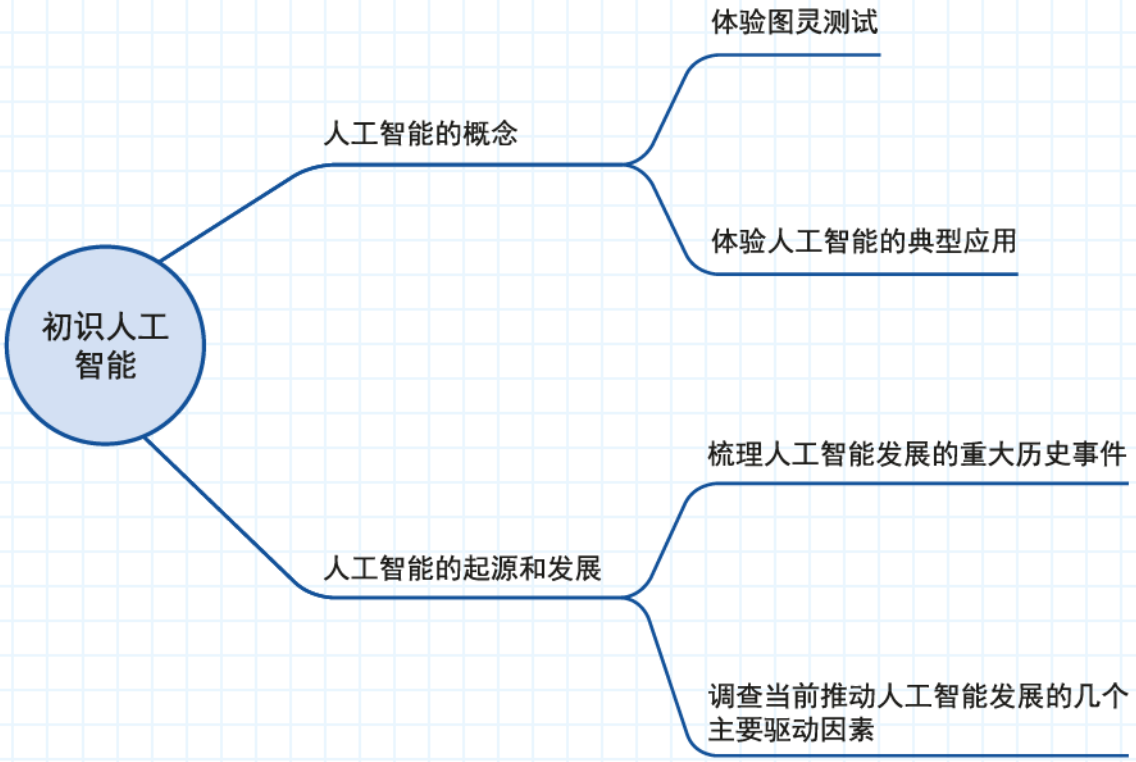
3. 人工智能是计算机科学的一个分支，它属于世界三大尖端技术之一。所谓三大尖端技术是指_____、_____和_____。

4. 1956年的_____会议是人工智能诞生的标志性事件，会上正式提出了人工智能的概念。

5. 支撑人工智能发展的主要驱动因素可以归纳为三点，即_____、_____和_____。

6. 人工智能处理器（芯片）主要有_____、_____和_____。

单元学习总结


初识人工智能
人工智能的概念
体验图灵测试
体验人工智能的典型应用
人工智能的起源和发展
梳理人工智能发展的重大历史事件
调查当前推动人工智能发展的几个主要驱动因素

第 2 单元　知识表示及其应用

通过前面的学习我们知道，科学家很早以前就提出了人工智能的概念，不同的学者对它有不同的解释，我们无法回避界定人工智能内涵的问题。人工智能到底包含哪些内容？或者说，人工智能到底可以做哪些“智能”的事情？人工智能学者尼尔斯·J.尼尔森（Nils. J. Nilsson）说：“广义地讲，人工智能是关于人造物的智能行为，而智能行为包括知觉、推理、学习、交流和在复杂环境中的行为。”由此可见，人工智能由“人造物”的诸多行为构成，而这些行为是可以被计算机理解和表示的。因此，作为人工智能的知识体系还应该回答如何采用计算机能够理解的符号系统描述智能行为的问题，即知识表示。人们希望计算机能够智能地搜索并进行逻辑推理，进而像人一样思考问题、解决问题，同时计算机还应该会学习，具备自我完善的能力。而搜索、推理、学习这一切能力都源于知识表示，可以说，知识表示是人工智能的基础。

在本单元中，我们将学习人工智能学科领域中最基本的内容，即知识表示的发展历程、知识的空间表示和逻辑表示，进而了解人工智能的搜索技术和逻辑推理技术。

本单元我们将通过“探究人工智能基础知识”项目，围绕人工智能中知识表示、智能搜索和逻辑推理的技术与应用来学习人工智能的经典内容。

为了完成该项目，需要考虑以下问题：什么是知识表示？什么是知识的状态—过程表示方法？什么是知识的逻辑表示？什么是基于状态空间的搜索？为此，我们需要完成以下任务：

- ◆ 通过认识典型知识表示方法了解知识表示的演化过程
- ◆ 通过解决汉诺塔问题认识知识的状态—过程表示方法
- ◆ 通过“将语言符号化”认识知识的逻辑表示方法
- ◆ 通过解决探路问题学习搜索技术

2.1 知识表示发展史

计算机科学领域一直使用数据、信息、知识和智慧这四个概念描述所处理的对象。其中，数据是最基本的，计算机能够处理的所有对象都可以称为数据；信息是“有用的”数据，信息论的创始人香农（Claude Elwood Shannon）说过，“信息是用来消除随机不定性的东西”；而知识是经过消减、塑造、解释、选择和转换的信息，它是由特定领域的描述、关系和过程组成的；智慧则是更高级的概念，此处不做深入探讨。

上述四个概念存在着隐藏的关系，即前一个概念是后一个概念的加工对象。试想一下，要想获得“信息”，就要加工“数据”；而要想获得“知识”，则要加工“信息”。人工智能也被称作“人工智慧”，要想获得“智慧”，我们就要加工“知识”。要加工“知识”，最基本的要求是解决知识的计算机表示问题，只有不同来源、不同形式的各种“知识”都可以用计算机表示，我们才可以对其进行加工，从而形成“人工智慧”，这正是我们学习知识表示及其应用的原因。

本节将围绕“知识表示的演化历程”，从知识表示发展的角度展开学习，通过“体验动物识别专家系统”“描述家庭关系”等应用案例，帮助同学们全面了解知识表示的各种方法，认识知识表示在人工智能领域中的重要性。

学习目标

★ 了解知识表示的发展历程。

★ 通过典型知识表示的应用，认识知识表示的重要性。

知识表示的发展主要经历了一阶谓词逻辑、产生式规则、状态空间、问题归约、框架、脚本、语义网络、语义网、知识图谱、事理图谱等几种不同方法的演化阶段，每种方法都有自己的知识表示形式、知识的组织结构以及求解问题的过程。总结这些知识表示方法可以看出，人工智能领域的研究者总是试图采用人对知识的认知途径模拟知识的表示。例如，用逻辑的方法表示知识是人类逻辑思维的习惯，于是在知识表示中就产生了一阶谓词逻辑，在后续的语义网、知识图谱和事理图谱中也都能看到逻辑的身影。人类还擅长形象思维，认为知识是普遍联系的，于是在知识表示中就产生了知识的空间表示方法，语义网络、语义网、知识图谱和事理图谱就是知识空间表示方法的典型代表。另外，人工智能中还试图模拟人类解决问题的思维方式，即问题状态的描述和问题状态的转换，于是在知识表示中就产生了状态—过程表示方法，状态空间、问题归约就属于此类知识表示方法。

本节我们重点了解知识表示发展过程中重要且具有广泛应用的知识表示方法，并通过实例体验它们的特点。

任务　通过认识典型知识表示方法了解知识表示的演化过程

※ 活动1　体验动物识别专家系统

专家系统是人工智能发展过程中的一个里程碑，自20世纪60年代开始，其应用产生了一定的经济效益和社会效益，成为人工智能领域中最活跃、最受重视的部分。专家系统通常由人机交互界面、规则库（知识库）、控制系统和推理机、综合数据库（存放已知条件）、知识获取机制五部分构成，如图2.1.1所示。专家系统的目标是模拟人类专家的推理思维过程，其模式一般是将专家的知识和经验，用一种知识表示方法存入计算机，系统对输入的事实进行推理，做出判断和决策。专家系统的知识表示方法有很多，最具代表性的是产生式规则。

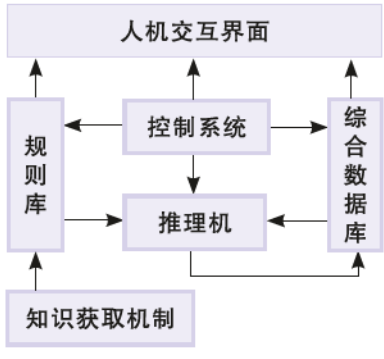


图 2.1.1　专家系统的基本结构

产生式规则

美国数学家波斯特（E.Post）在1943年提出用符号语言构造产生式计算模型“<前件> → <后件>”的方法。他指出任何数学系统、逻辑系统都可视为一个产生式集合，即规定了如何将一个符号串变换成另一个符号串。图2.1.2给出了产生式规则知识表示方法所包含的内容。

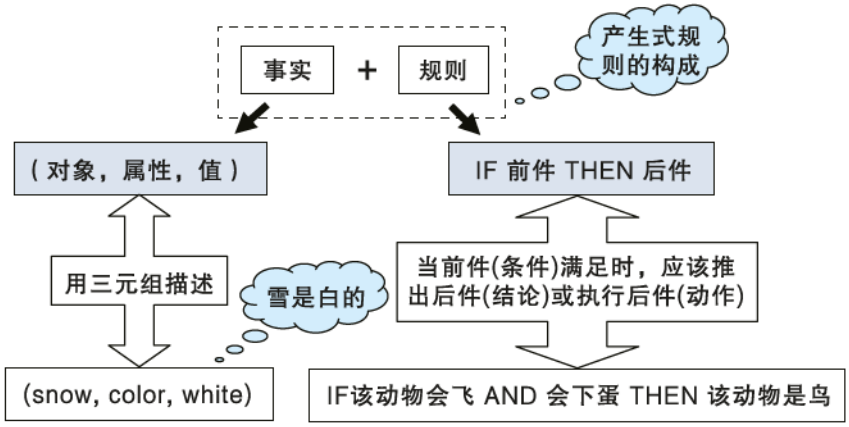


图2.1.2 产生式规则的构成说明

例如，我们要构建一个动物识别专家系统，首先要构建由用于识别动物的已知知识（通常是专家提供）组成的产生式规则，如表2.1.1所示。

表2.1.1 用于识别动物的产生式规则

编号	含义	产生式规则
r1	有毛发——哺乳动物	IF 该动物有毛发 THEN 该动物是哺乳动物
r2	有奶——哺乳动物	IF 该动物有奶 THEN 该动物是哺乳动物
r3	有羽毛——鸟	IF 该动物有羽毛 THEN 该动物是鸟
r4	会飞，会下蛋——鸟	IF 该动物会飞 AND 会下蛋 THEN 该动物是鸟

续表

编号	含义	产生式规则
r5	吃肉——食肉动物	IF 该动物吃肉 THEN 该动物是食肉动物
r6	有犬齿，有爪，眼盯前方——食肉动物	IF 该动物有犬齿 AND 有爪 AND 眼盯前方 THEN 该动物是食肉动物
r7	哺乳动物，有蹄——有蹄类动物	IF 该动物是哺乳动物 AND 有蹄 THEN 该动物是有蹄类动物
r8	哺乳动物，反刍动物——有蹄类动物	IF 该动物是哺乳动物 AND 是反刍动物 THEN 该动物是有蹄类动物
r9	哺乳动物，食肉动物，黄褐色，暗斑点——金钱豹	IF 该动物是哺乳动物 AND 是食肉动物 AND 是黄褐色 AND 身上有暗斑点 THEN 该动物是金钱豹
r10	哺乳动物，食肉动物，黄褐色，黑色条纹——虎	IF 该动物是哺乳动物 AND 是食肉动物 AND 是黄褐色 AND 身上有黑色条纹 THEN 该动物是虎
r11	有蹄类动物，长脖子，长腿，暗斑点——长颈鹿	IF 该动物是有蹄类动物 AND 有长脖子 AND 有长腿 AND 身上有暗斑点 THEN 该动物是长颈鹿
r12	有蹄类动物，黑色条纹——斑马	IF 该动物是有蹄类动物 AND 身上有黑色条纹 THEN 该动物是斑马
r13	鸟，长脖子，长腿，不会飞，黑白二色——鸵鸟	IF 该动物是鸟 AND 有长脖子 AND 有长腿 AND 不会飞 AND 有黑白二色 THEN 该动物是鸵鸟
r14	鸟，会游泳，不会飞，黑白二色——企鹅	IF 该动物是鸟 AND 会游泳 AND 不会飞 AND 有黑白二色 THEN 该动物是企鹅
r15	鸟，善飞——信天翁	IF 该动物是鸟 AND 善飞 THEN 该动物是信天翁

产生式规则专家系统提供了按照产生式规则进行推理的机制。动物识别专家系统的产生式规则共有15条，可以识别的动物有虎、金钱豹、斑马、长颈鹿、企鹅、信天翁、鸵鸟，共7种。规则中与这7种动物相对应的被称为目标规则，不是目标规则的被称为中间规则。这15条规则存放在规则库中。

假如通过观察获取的某种动物的特征是该动物身上有暗斑点，有长脖子，有长腿，有奶，有蹄，这是什么动物呢？已知条件也被称为事实，事实首先被放入综合数据库中。图2.1.3给出了推理过程。

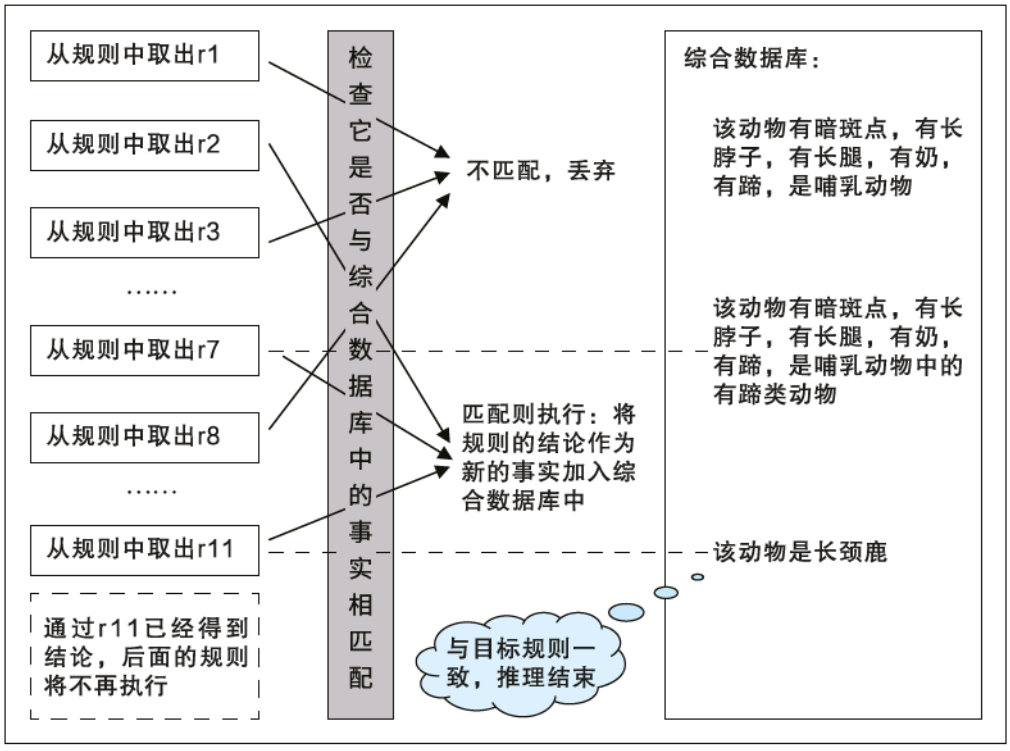


图 2.1.3　产生式规则专家系统的推理实例

通过上面的介绍，我们已经对产生式规则以及产生式规则专家系统有了初步的认识。由于采用了计算机便于理解的知识表示方法，产生式规则专家系统可以方便地用计算机程序实现。教科书配套资源中给出了用Python语言编写的基于产生式规则的动物识别专家系统的源代码。它的运行界面如图2.1.4和图2.1.5所示。结合所学知识，完成以下活动内容。

◆ 在教科书配套资源中找到该源代码，在Python环境中运行。

◆ 选择已知动物的特征或类型，验证你对产生式规则的理解，体验产生式规则专家系统的特点。

◆ 思考并回答表2.1.2中的问题。

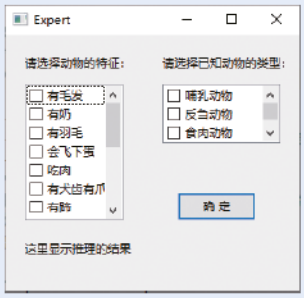


图 2.1.4　动物识别专家系统初始界面

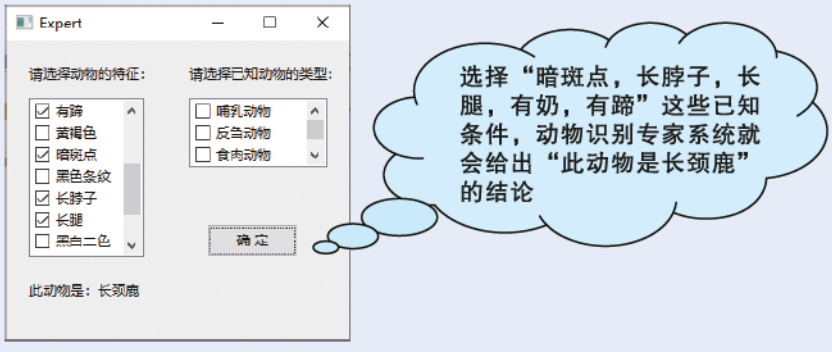


图 2.1.5　动物识别专家系统的推理结果

表 2.1.2　体验Python语言构建的动物识别专家系统活动记录表

序号	主题	结果记录
1	设计程序时，需要将用户输入的文字与规则库中的规则前件进行匹配，说说本程序是如何实现匹配的	
2	一次匹配得到的结论可能只是“中间假设”，程序需要一直匹配下去，直到结论是目标集合中的一个结论（有解）或者所有规则都使用过但其结论仍不在目标集合中（无解），说说本程序是如何实现多次匹配的	
3	本程序使用了Python可视化程序设计技术，说说它使用了哪种Python图形包，以及它的可视化设计是如何实现的，即在本程序中使用了哪些语句	

※ 活动2　描述家庭关系

知识表示“家族”中一个重要的分支是知识的空间表示方法，语义网络、语义网、知识图谱和事理图谱都是这个家族的成员。知识的空间表示方法类似人类的形象思维，它们都是用事物和事物之间的联系描述知识。目前，知识图谱和事理图谱是人工智能研究领域的两大热点，代表了人工智能技术发展的方向。

本活动中，我们将初步接触知识的各种空间表示方法，并通过实例了解这类方法的特点。

● 知识的空间表示方法演化

知识的空间表示方法经历了从语义网络到语义网，再到知识图谱，直到事理图谱的演化过程，图2.1.6给出了它们开始的年代和各种表示方法的简单解释。其中，本体是专业领域的术语集，并不是知识表示方法，但是很多知识表示方法都和本体有关；而当代各种知识表示方法都离不开万维网，因此，在图中也标注了本体和万维网两个词条。

语义网络是早期的知识表示方法，它只强调表示方法，没有一套计算机实现的完整标准。这种情况下，使用者可以随意用不同的计算机语言、不同的实现方式实现语义网络，造成了语义网络表达的混乱。语义网的提出在万维网概念提出之后，它用于指代一类技术标准。在万维网诞生之初，网络上的内容只是人类可读，而计算机无法理解和处理。比如，在浏览一个网页时，人能够理解网页上的内容，而计算机只知道这

是一个网页，并不清楚网页中写的是什么，也不清楚链接指向的页面和当前页面有何关系。语义网是为了使网络上的数据变得机器可读而提出的一个通用框架。Semantic指的是用更丰富的方式来表达数据背后的含义，让机器能够理解数据。Web是希望这些数据相互链接，组成一个庞大的信息网络，像互联网中相互链接的网页一样，只不过基本单位变为粒度更小的数据。

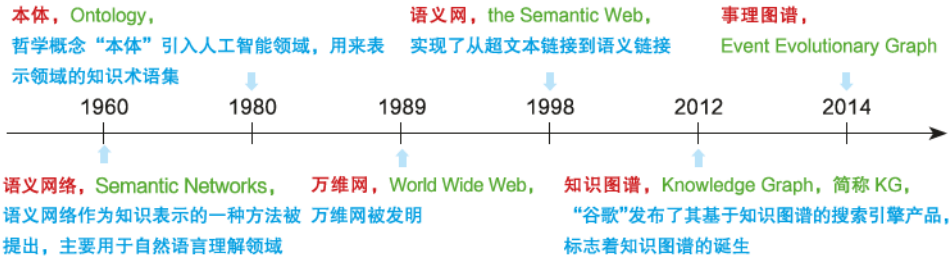


图 2.1.6 知识的空间表示方法演化图

语义网也叫语义Web，它和语义网络只有一字之差，但它们是不同的表示方法。语义网络是早期的知识表示方法，它只强调表示方法，语义网络是早期的概念，描述场景或概念之间的联系，它对描述的知识内容没有要求。知识的本体描述出现在语义网络之后，它对描述的知识内容提出了要求，即本体描述的是面向特定领域公认的概念，是规范的、明确的、形式化的、共享的和公认的，本体的开发需要人类专家的直接参与。无论是语义网络还是本体都没有强调实现方法，而知识图谱则是产品级的知识表示体系，它有一整套知识抽取、融合、挖掘和评估的方法，可以说知识图谱是语义网络和本体的具体实现。

语义网络

为了深入了解知识的空间表示方法，我们有必要从空间表示的源头——语义网络开始，学习和认识这种表示方法的结构和特点。语义网络是心理学家奎廉（J.R.Quillian）于1968年在研究人类联想记忆时提出的一种心理学模型，他认为记忆是由概念间的联系实现的。随后，奎廉又把它用作知识表示。语义网络是一种用实体及其语义关系来表达知识的有向图。其中，节点代表实体，表示各种事物、概念、情况、属性、状态、事件、动作等；弧代表语义关系，表示它所连接的两个实体之间的语义联系，它必须带有标识。语义网络中最基本的语义单元称为语义基元，可用三元组表示为（节点1，弧，节点2）。图2.1.7是一个语义网络的实例。

有向图

图（Graph）是数学中表示物件与物件之间关系的方法，是图论的基本研究对象。一个图看起来是由一些小圆点（叫作顶点或节点）和连接这些圆点的线（叫作弧或边）组成的。有向图是一个二元组 $<V, E>$，其中 V 是非空集合，称为顶点集；E 是 $V \times V$ 的子集，称为弧集。直观地说，若图中的每条边都是有方向的，则称为有向图。有向图中的边是由两个顶点组成的有序对，有序对通常用尖括号表示，如 $<v_i, v_j>$ 表示一条有向边，其中 v_i 是边的始点，v_j 是边的终点。$<v_i, v_j>$ 和 $<v_j, v_i>$ 代表两条不同的有向边。

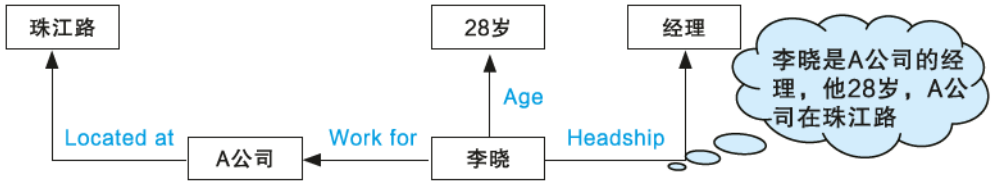


图 2.1.7　一个语义网络的实例

知识图谱与事理图谱

知识图谱（Knowledge Graph）出现于2012年前后，其目标是致力于改善网络搜索结果，描述真实世界中存在的各种实体和概念，以及这些实体、概念之间的关联关系。从2012年开始，国内外互联网企业纷纷构建了自己的知识图谱，如微软的Probase、搜狗的知立方、百度知识图谱等。知识图谱在语义搜索、智能问答、数据挖掘、数字图书馆、推荐系统等领域有着广泛的应用。

关于知识图谱概念，目前没有标准的定义，比较权威的说法是：知识图谱是由一些相互联结的实体和它们的属性构成的，可以将知识图谱看成一张巨大的图，图中的节点表示实体或概念，而图中的边则构成关系。

在知识图谱中，每个实体和概念都使用一个全局唯一确定的ID来标识，这个ID对应目标的标识符（identifier），这种做法与一个网页和一个URL对应相类似。知识图谱中的实体之间存在关联关系。实体可以拥有属性，用于刻画实体的内在特性，每个属性都是以“<属性，属性值>对（Attribute-Value Pair，AVP）”的方式来表示。图2.1.8给出了一个知识图谱的实例。

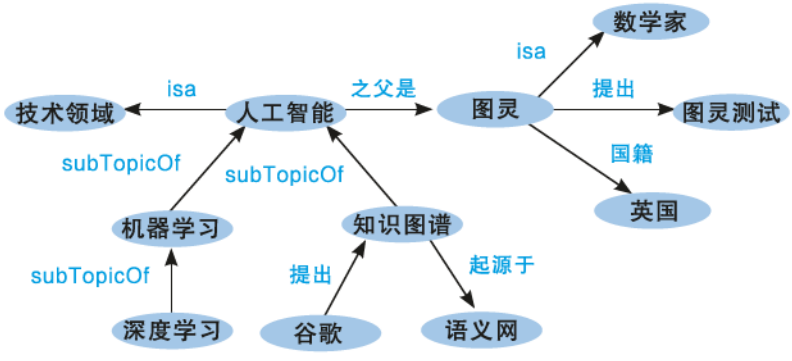


图 2.1.8　一个知识图谱的实例

事理图谱是知识图谱的发展。知识图谱表达的是概念与概念的关系，缺乏对“过程逻辑”的描述，而事理图谱弥补了这一缺陷。事理图谱也是一个有向图，它的节点代表事件，有向边代表事件之间的顺承、因果关系。图2.1.9给出了一个事理图谱的实例。

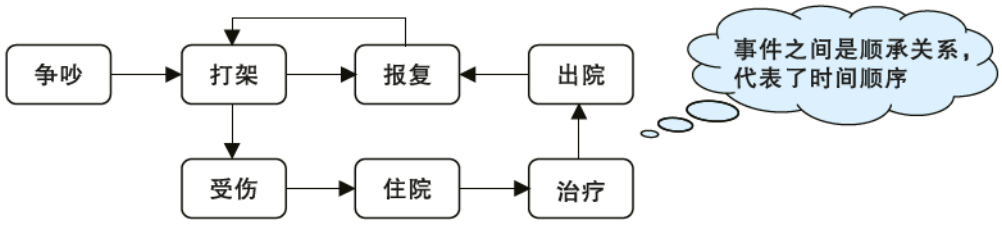


图 2.1.9 一个事理图谱的实例

通过前面的学习，我们已经认识了语义网络、知识图谱和事理图谱，为了进一步巩固知识，请完成以下活动。

构建一个“我的家庭”语义网络，填写表2.1.3。要求：（1）包括三个或三个以上家庭成员；（2）对于每个家庭成员，需要描述他（她）的基本情况，例如年龄、身高、体重等，还要描述他（她）的社会身份，例如从事何种工作、具体的工作单位、从何时开始工作、何时退休等；（3）描述家庭成员之间的关系。

表 2.1.3 “我的家庭”语义网络结果记录表

结果记录

知识图谱的一个重要应用是对搜索引擎的支持。例如，百度的知识图谱搜索引擎被称为“百度知心”，搜狗的知识图谱搜索引擎被称为“知立方”。有了知识图谱的支撑，搜索引擎变得更加强大。访问百度知识图谱的主页，单击“直接给出答案”专栏右侧的超链接“立即体验”，进入百度搜索引擎主页。在搜索栏中输入“妈妈爸爸的爸爸是谁？”，百度搜索引擎给我们提供的答案如图2.1.10所示。它能够回答的问题的复杂性对于传统的搜索引擎来说是难以想象的。

图 2.1.10 百度知识图谱搜索引擎的应用

针对表 2.1.4 中的问题，比较百度和搜狗两个搜索引擎提供的答案，看看哪一个更准确地回答了你的问题，并将结果记录到表 2.1.4 中。

表 2.1.4　网络搜索比较记录表

序号	问题	百度的解答	搜狗的解答
1	家庭的主要成员有哪些		
2	爸爸的爸爸的爸爸的爸爸是谁		
3	女儿如何称呼爸爸的哥哥		

知识图谱在移动设备上也有着广泛的应用。进入“微信—发现—小程序”，打开其中的“腾讯AI体验中心”，选择“自然语言处理”面板，其中的最后一项“知识问答”是知识图谱支撑下的智能问答系统。为了“了解家庭关系”，请同学们设计与家庭有关的问题，然后进入“腾讯AI体验中心”，与“知识问答系统”展开对话，并将结果记录到表2.1.5中。

表 2.1.5　用移动设备的知识问答系统“了解家庭关系”活动记录表

序号	与家庭关系有关的问题	“知识问答系统”给出的解答
1		
2		
3		
4		
5		
6		

2.2 知识的状态—过程表示

上一节中提到，人工智能试图模拟人类解决问题的思维方式。研究表明，人类解决问题的过程包括问题状态的描述和问题状态的转换。在知识表示的各种方法中，模拟人类解决问题的思维方式的一类方法被称为状态—过程表示方法，具体主要包括状态空间方法和问题归约方法。

本节我们将围绕“解决汉诺塔问题”展开学习，通过分别使用状态空间方法和问题归约方法解决汉诺塔问题，帮助同学们深入理解这两种方法的内涵。

学习目标

★ 理解状态空间方法的原理，明确状态空间方法解决问题的过程。

★ 理解问题归约方法的原理，明确问题归约方法解决问题的过程。

人类求解问题时，首先为问题选择适当的状态及操作的形式化描述方法；然后从某个初始状态出发，每次使用一个操作，递增地建立操作序列，直到达到目标状态为止；由初始状态到目标状态所使用的算符序列就是该问题的一个解。人类求解问题的例子如图2.2.1所示。模拟人求解问题的思路，人工智能技术也有类似的方法，它们是状态空间方法和问题归约方法。利用状态空间方法求解问题时，用状态空间形式化描述知识，而利用问题归约方法求解问题时，用与/或树表示知识，它们都是知识的状态—过程表示方法。

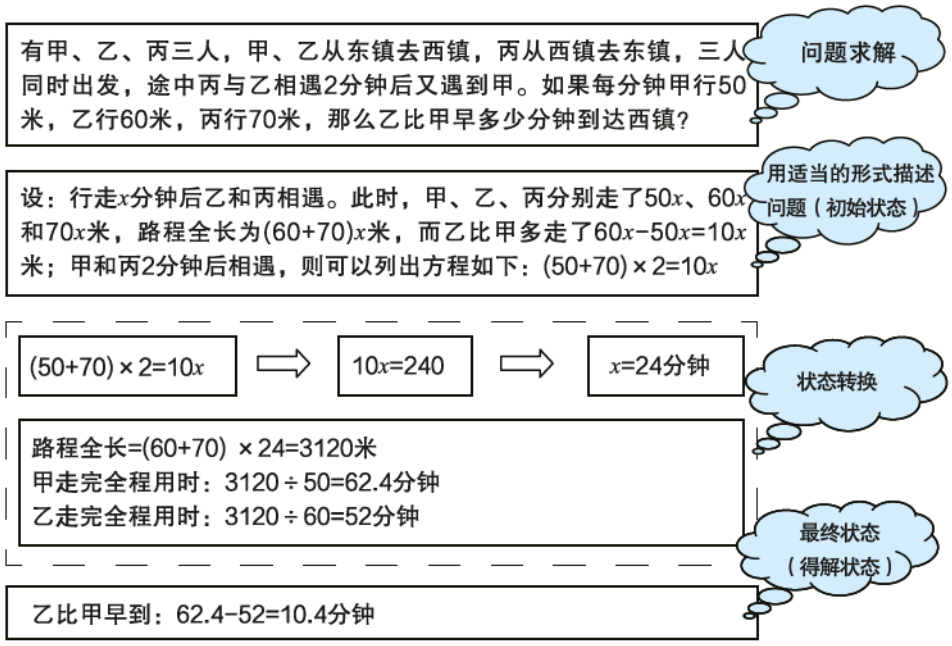


图 2.2.1　人类求解问题的过程描述

图2.2.2展示的就是著名的汉诺塔问题。接下来我们将用状态空间和与/或树表示汉诺塔问题的各种状态，并分别使用状态空间方法和问题归约方法求解汉诺塔问题。通过本节的学习，同学们可以充分理解知识的状态—过程表示方法，逐步养成描述问题状态、通过状态转换求解问题的思维习惯。

在印度北部的圣庙里，一块黄铜板上插着三根宝石针，在其中一根宝石针上从下到上地穿好了由大到小的64片金片，这就是所谓的汉诺塔。不论白天黑夜，总有一个僧侣在按照下面的法则移动这些金片：一次只移动一片，不管在哪根宝石针上，小片必须在大片上面。僧侣们预言，当所有的64片金片都从那根宝石针上移到另外一根宝石针上时，世界就将在一声霹雳中消失。

图 2.2.2　汉诺塔问题

任务　通过解决汉诺塔问题认识知识的状态—过程表示方法

※ 活动1　用状态空间方法解决问题

下面我们来研究著名的汉诺塔问题，为了降低难度，这里我们只研究二阶汉诺塔问题，这时只有*A*和*B*两片“金片”。

求解问题时的状态如图2.2.3所示。

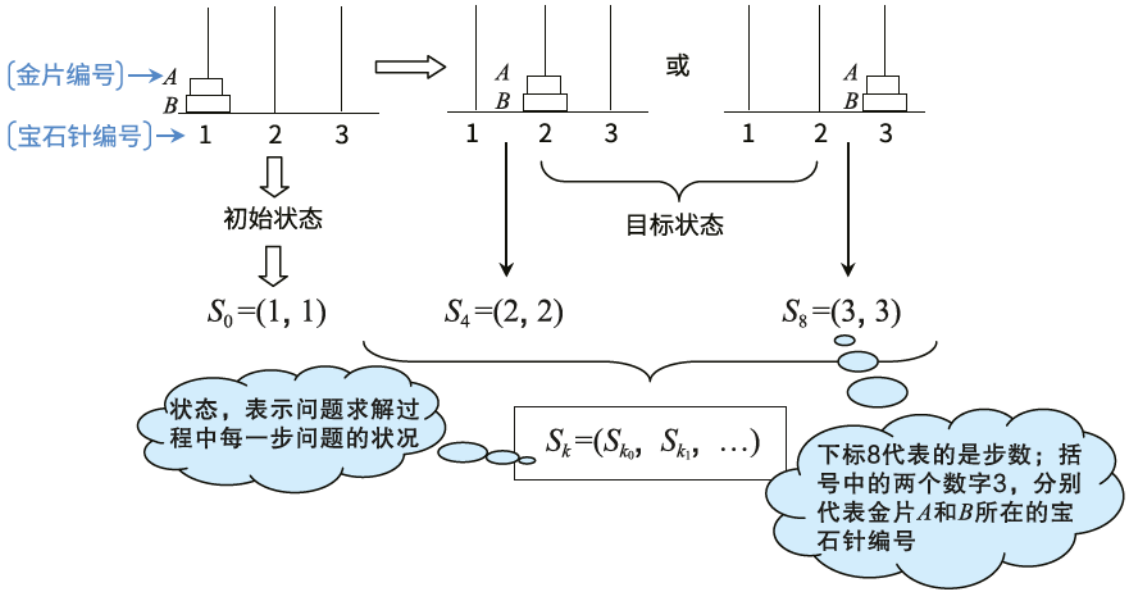


图 2.2.3　求解问题时的状态

为了完成移动，必须要“操作”。我们这样定义“操作”：把问题从一种状态变换为另一种状态的手段。求解问题时的操作如图2.2.4所示。

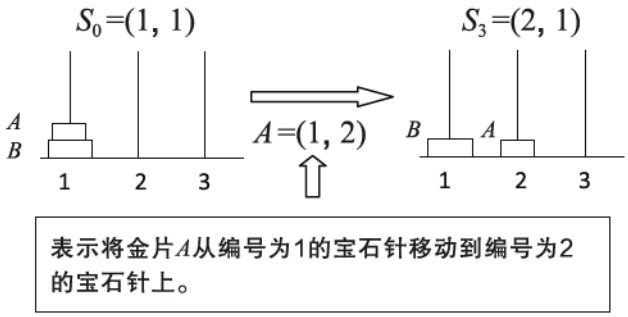


图 2.2.4　求解问题时的操作

状态空间用来描述一个问题的全部状态以及这些状态之间的相互关系，常用三元组（S，F，G）表示。其中，S为问题的所有初始状态的集合；F为操作的集合；G为目标状态的集合。汉诺塔问题的状态空间如图2.2.5所示。

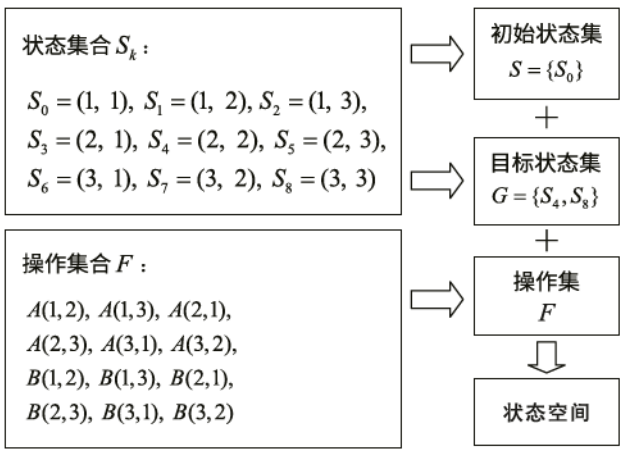


图 2.2.5　汉诺塔问题的状态空间

状态空间也可用一个赋值的有向图来表示，该有向图称为状态空间图。在状态空间图中，节点表示问题的状态，有向边表示操作。根据上述9种可能的状态和12种操作，可构成二阶汉诺塔问题的状态空间图，如图2.2.6所示。从初始节点(1, 1)到目标节点(2, 2)及(3, 3)的任何一条路径都是问题的一个解。其中，最短的路径长度是3，它由3个操作组成。例如，从(1, 1)开始，通过使用操作$A(1, 3)$、$B(1, 2)$及$A(3, 2)$，可到达(2, 2)。

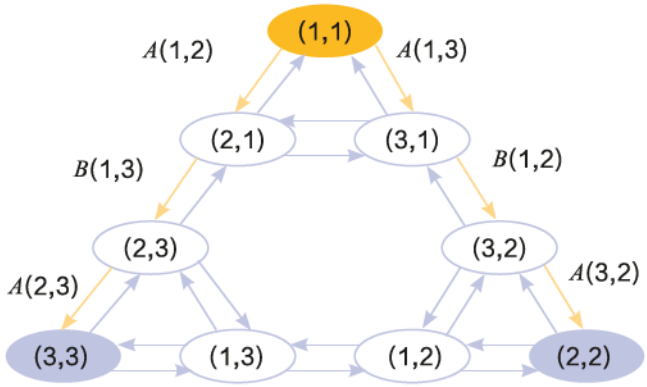


图 2.2.6　二阶汉诺塔问题的状态空间图

通过上面的介绍，我们对状态空间知识表示和用状态空间方法解决问题已经有了初步的认识，下面将开展活动，进一步巩固所学知识。

◆ **用状态空间方法解决猴子摘香蕉问题。**

问题描述：

如图2.2.7所示，在一个房间里，天花板上挂着一串香蕉，有一只猴子可在房间里任意活动（到处走动、推移箱子、攀登箱子等）。设房间里有一个可被猴子移动的箱子，且猴子登上箱子时才能摘到香蕉，问猴子在某一状态下（设猴子位置为a，箱子位置为c，香蕉位置在b）如何行动可摘到香蕉。

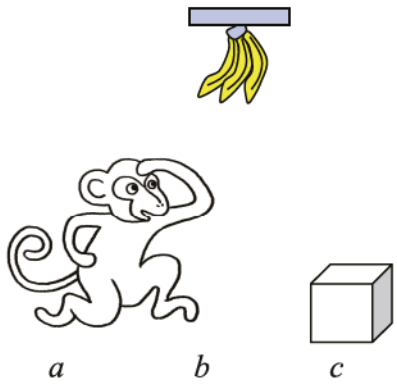


图 2.2.7　猴子摘香蕉问题

表2.2.1 解决猴子摘香蕉问题活动记录表1

项目	方案	说明
如何描述问题的状态		
如何描述问题的操作		
状态空间描述		
状态空间图		

※ 活动2 用问题归约方法解决问题

我们进一步用问题归约方法研究汉诺塔问题。当一个问题比较复杂时，可通过分解或变换，将其转化为一系列较简单的子问题，然后通过对这些子问题的求解来实现对原问题的求解，这就是问题归约方法的基本思想。在用问题归约方法解决汉诺塔问题之前，我们先了解一下它用到的知识表示方法。

如果一个问题P可以归约为一组子问题P_1，P_2，…，P_n，并且只有当所有子问题P_i都有解时原问题P才有解，任何一个子问题P_i无解都会导致原问题P无解，则称此种归约为问题的分解，如图2.2.8所示。

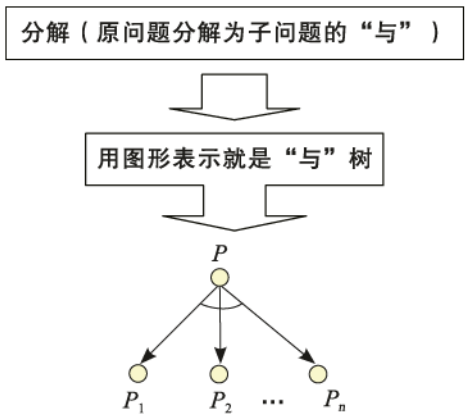


图2.2.8 分解和“与”树

如果一个问题P可以归约为一组子问题P_1，P_2，…，P_n，并且子问题P_i中只要有一个有解则原问题P就有解，只有当所有子问题P_i都无解时原问题P才无解，则称此种归约为问题的等价变换，简称变换，如图2.2.9所示。

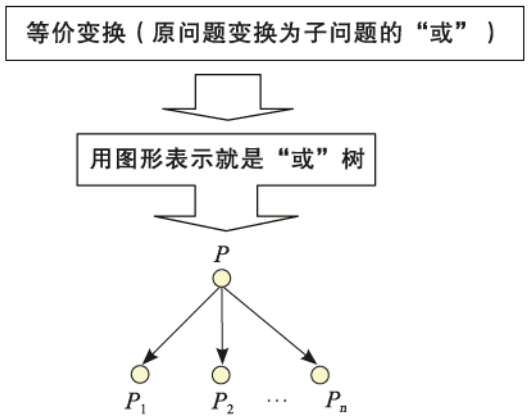


图 2.2.9　等价变换和“或”树

在解决问题时，“与”树和“或”树往往是结合在一起使用的，就像图2.2.10所示的那样，我们称它为“与/或树”，它是用状态—过程方法表示知识的另一种形式。在与/或树中，没有子节点的节点称为端节点，节点P_{33}就是端节点。本原问题所对应的节点称为终止节点。所谓本原问题就是不可或不需再通过变换化简的“原子”问题，它是不用证明而自然成立的，如公理、已知事实或已证明的问题等。可见，终止节点一定是端节点，但端节点却不一定是终止节点。

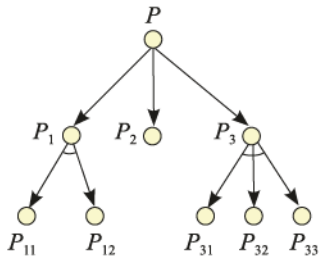


图 2.2.10　与/或树

解树

在与/或树中，满足以下三个条件之一的节点为可解节点。

（1）任何终止节点都是可解节点；

（2）对“或”节点，当其子节点中至少有一个为可解节点时，则该或节点就是可解节点；

（3）对“与”节点，只有当其子节点全部为可解节点时，该与节点才是可解节点。

在与/或树中，满足以下三个条件之一的节点为不可解节点。

（1）不为终止节点的端节点是不可解节点；

（2）对“或”节点，若其全部子节点都为不可解节点，则该或节点是不可解节点；

（3）对“与”节点，只要其子节点中有一个为不可解节点，则该与节点是不可解节点。

由可解节点构成，并且通过这些可解节点可以推出初始节点（对应着原始问题）是可解节点的子树，被称为解树。在解树中一定包含初始节点。

在如图2.2.11所示的与/或树中，用红线表示的子树就是一个解树。节点*P*为原始问题节点，节点*t*是终止节点。根据可解节点的定义，很容易推出原始问题*P*为可解节点。问题归约求解过程实际上就是生成解树，即证明原始节点是可解节点的过程。

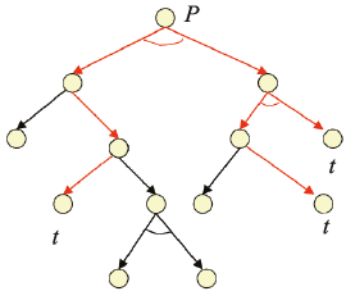


图 2.2.11　与/或树中的解树

下面用问题归约方法求解三阶（有三个“金片”的情况）汉诺塔问题，其要求如图2.2.12所示。

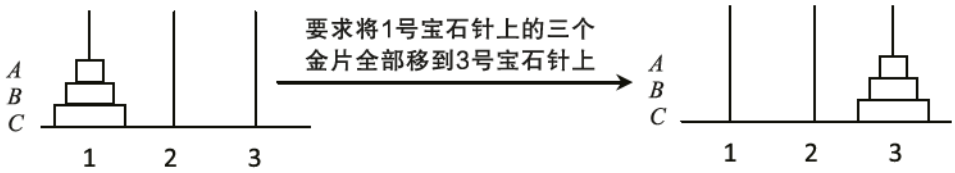


图 2.2.12　三阶汉诺塔问题

三阶汉诺塔问题的状态描述：

用三元组(*i*, *j*, *k*) 表示问题在任一时刻的状态，其中*i* 代表*C*所在的宝石针编号，*j* 代表*B*所在的宝石针编号，*k* 代表*A*所在的宝石针编号。

问题分析如下。

用“→”表示状态的转换。

利用问题归约方法，原问题可分解为以下三个子问题。

（1）把金片 *A* 及 *B* 移到 2 号宝石针上的双金片移动问题，即 (1, 1, 1) → (1, 2, 2)；

（2）把金片 *C* 移到 3 号宝石针上的单金片移动问题，即 (1, 2, 2) → (3, 2, 2)；

（3）把金片 *A* 及 *B* 移到 3 号宝石针上的双金片移动问题，即 (3, 2, 2) → (3, 3, 3)。

子问题（1）和（3）都是一个二阶汉诺塔问题，它们都还可以再继续进行分解；子问题（2）是本原问题，已不需要再分解。

三阶汉诺塔问题的与/或树如图2.2.13所示。

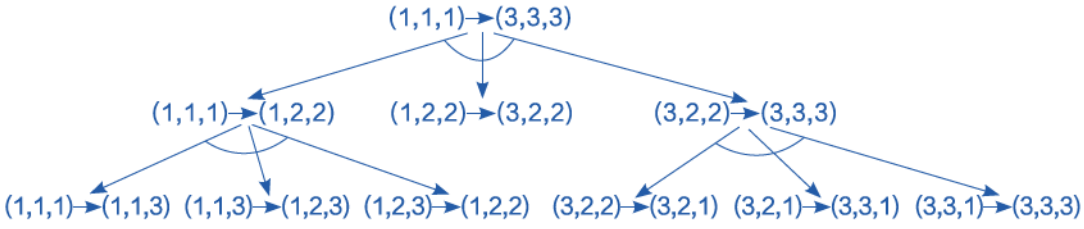


图 2.2.13　三阶汉诺塔问题的与/或树

结论：

在该与/或树中，有7个终止节点，它们分别对应着7个本原问题。如

果把这些本原问题从左至右排列起来，即可得到原始问题的解：

(1, 1, 1)→(1, 1, 3)，(1, 1, 3)→(1, 2, 3)，(1, 2, 3)→(1, 2, 2)，(1, 2, 2)→(3, 2, 2)，(3, 2, 2)→(3, 2, 1)，(3, 2, 1)→(3, 3, 1)，(3, 3, 1)→(3, 3, 3)

通过上面的介绍，我们对与/或树知识表示和用问题归约方法解决问题已经有了初步的认识，下面将开展活动，进一步巩固所学知识。

◆ **用问题归约方法解决猴子摘香蕉问题。**

请采用问题归约方法解决猴子摘香蕉问题，填写表2.2.2。

表 2.2.2　解决猴子摘香蕉问题活动记录表2

项目	方案	说明
状态描述		
问题分析		
与/或树表示		
结论展示		

2.3 知识的逻辑表示

福尔摩斯通过基本演绎法还原案件真相的本领令我们记忆犹新。所谓基本演绎法就是逻辑推理，它是智能最完美的体现。人工智能兴起之初，科学家就致力于研究逻辑的计算机表示方法，力求使计算机拥有推理能力。逻辑最早来源于古希腊的论辩，后来数学家希望通过符号形式化地表示逻辑，于是诞生了数理逻辑。计算机中表示逻辑的常用方法称为一阶谓词逻辑，它是数理逻辑的子集。智能推理是指以逻辑表示为基础，把人类的思维活动形式化、符号化，使其能在计算机上实现的过程。

本节我们将围绕“将语言符号化”任务展开学习，通过使用一阶谓词逻辑进行语言判断和过程描述两个角度的活动，帮助同学们深入认识知识的逻辑表示方法。

学习目标

★ 认识一阶谓词逻辑表示法中的基本概念。

★ 了解使用一阶谓词逻辑进行语言判断和过程描述的基本方法。

任务 通过“将语言符号化”认识知识的逻辑表示方法

※ 活动1 用一阶谓词逻辑表示法表示命题知识

● 一阶谓词逻辑

知识的一阶谓词逻辑表示是计算机中对自然语言进行形式化处理的基础，是计算机理解自然语言的重要步骤，也是计算机能够完成推理等高级智能活动的技术保障。一阶谓词逻辑框架是由一些术语、符号和符号运算后的真值组成的。

一阶谓词逻辑中包括以下一系列术语。

论域：由所讨论对象的全体构成的集合，亦称为个体域。

个体：论域中的元素。

断言：一个陈述句称为一个断言，断言是知识在自然语言中的表示形式。

命题：具有真假意义的断言称为命题，命题也被称为命题知识。

真值：命题的真或假的结论被称为真值，它只有T和F两个值，也就是“真”和“假”。

谓词：在谓词逻辑中，命题是用形如$P(x_1, x_2, \cdots, x_n)$的谓词来表示的。其中，P是谓词名，是命题的谓语，表示个体的性质、状态或个体之间的关系；$x_1, x_2, \cdots, x_n$为个体，也叫变量，它们是命题的主语，表示独立存在的事物或概念。

一阶谓词逻辑中还包含连词和量词。其中，连词有以下几个。

$\neg$：“非”或者“否定”，表示对其后面的命题的否定。

$\vee$：“析取”，表示所连接的两个命题之间具有“或”的关系。

$\wedge$：“合取”，表示所连接的两个命题之间具有“与”的关系。

$\rightarrow$：“条件”或“蕴含”，表示“若……则……”的语义，$P \rightarrow Q$读作“如果P，则Q”。其中，P称为条件的前件，Q称为条件的后件。

$\leftrightarrow$：“双条件”，它表示“当且仅当”的语义，$P \leftrightarrow Q$读作“P当且仅当Q”。

量词有以下两个。

$\forall$：全称量词，表示“所有的”“任一个”“凡是”。

$\exists$：存在量词，表示“至少有一个”“存在有”。

一阶谓词逻辑中连词的真值如表2.3.1所示。

表 2.3.1　连词的真值表

P	Q	$\neg P$	$P \vee Q$	$P \wedge Q$	$P \rightarrow Q$	$P \leftrightarrow Q$
T	T	F	T	T	T	T
T	F	F	T	F	F	F
F	T	T	T	F	T	F
F	F	T	F	F	T	T

一阶谓词逻辑中量词真值的判断如下：

命题$(\forall x)P(x)$为真，当且仅当对论域中的所有x，都有$P(x)$为真；命题$(\forall x)P(x)$为假，当且仅当至少存在一个$x_i \in D$，使得$P(x_i)$为假；

命题$(\exists x)P(x)$为真，当且仅当至少存在一个$x_i \in D$，使得$P(x_i)$为真；命题$(\exists x)P(x)$为假，当且仅当对论域中的所有x，都有$P(x)$为假。

用一阶谓词逻辑表示知识的步骤如下：

（1）先根据要表示的知识定义谓词；

（2）再用连词、量词把这些谓词连接起来。

用一阶谓词逻辑表示“所有教师都有自己的学生”的过程如图2.3.1所示。

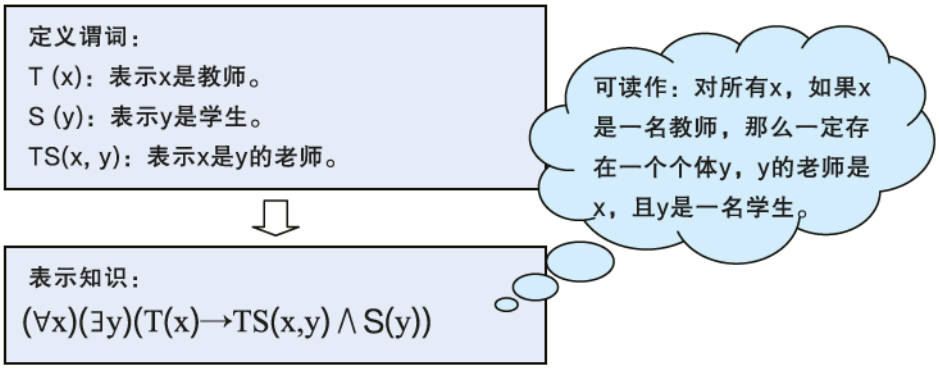


图2.3.1 用一阶谓词逻辑表示知识的实例

通过前面的学习，我们已经初步了解了一阶谓词逻辑的基本内容，接触了使用一阶谓词逻辑表示知识的实例，下面将开展活动，巩固所学知识。

◆ **用一阶谓词逻辑的“符号系统”表示下列命题，并将结果填入表2.3.2中。**

表2.3.2 用一阶谓词逻辑表示法表示命题知识的结果记录表

序号	主题	结果记录
1	王宏是信息班学生。	
2	王宏和李明是同班同学。	
3	凡是信息班的学生都喜欢编程序。	
4	有的人喜欢梅花，有的人喜欢菊花，有的人既喜欢梅花又喜欢菊花。	
5	不存在最大的整数。	
6	所有的整数不是偶数就是奇数。	

※ 活动2　用一阶谓词逻辑表示法描述过程

如图2.3.2所示，a、b、c是三个位置，c位置有一个机器人，a和b位置是两张桌子，a桌子上有一个盒子。机器人的任务是把这个盒子移动到b桌子上。如何表示机器人移动盒子的过程呢?

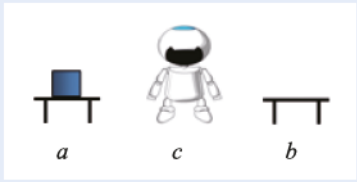


图 2.3.2　机器人移盒子

第一步，设计描述状态的谓词。不要忘记描述变量的个体域，即确定每个变量都可以取哪些值。结果如图2.3.3所示。

TABLE(x)：x是桌子
EMPTY(y)：y手中是空的
AT(y, z)：y在z处
HOLDS(y, w)：y拿着w
ON(w, x)：w在x桌面上

⇦

x的个体域是{a, b}
y的个体域是{robot}
z的个体域是{a, b, c}
w的个体域是{box}

图 2.3.3　设计描述状态的谓词

第二步，确定问题的初始状态和目标状态，如图2.3.4所示。

问题的初始状态：
AT(robot, c)
EMPTY(robot)
ON(box, a)
TABLE(a)
TABLE(b)

机器人行动的目标就是把问题的初始状态转换为目标状态，而要实现问题状态的转换则需要完成一系列操作

⇨

问题的目标状态：
AT(robot, c)
EMPTY(robot)
ON(box, b)
TABLE(a)
TABLE(b)

图 2.3.4　问题的初始状态和目标状态

第三步，设计操作谓词。

Goto(x, y)：从x处走到y处

Pickup(x)：在x处拿起盒子

Setdown(y)：在y处放下盒子

这些都是“复合”动作，计算机“不懂”其含义，它只懂“删除”和“添加”两个标准动作，故需要用删除和添加“解释”这些“复合”动作。

Goto(x, y)：前提状态是AT(robot, x)；动作1，删除谓词 AT(robot, x)，动作2，添加谓词AT(robot, y)。

Pickup(x)：前提状态是ON(box, x)，TABLE(x)，AT(robot, x)，EMPTY(robot)；动作1，删除谓词EMPTY(robot)，ON(box, x)；动作2，添加谓词HOLDS(robot, box)。

Setdown(x)：前提状态是AT(robot, x)，TABLE(x)，HOLDS(robot, box)；动作1，删除谓词HOLDS(robot, box)；动作2，添加谓词EMPTY(robot)，ON(box, x)。

至此，我们已经为用一阶谓词逻辑表示法描述过程做好了准备。图2.3.5表示的就是机器人移盒子的过程。

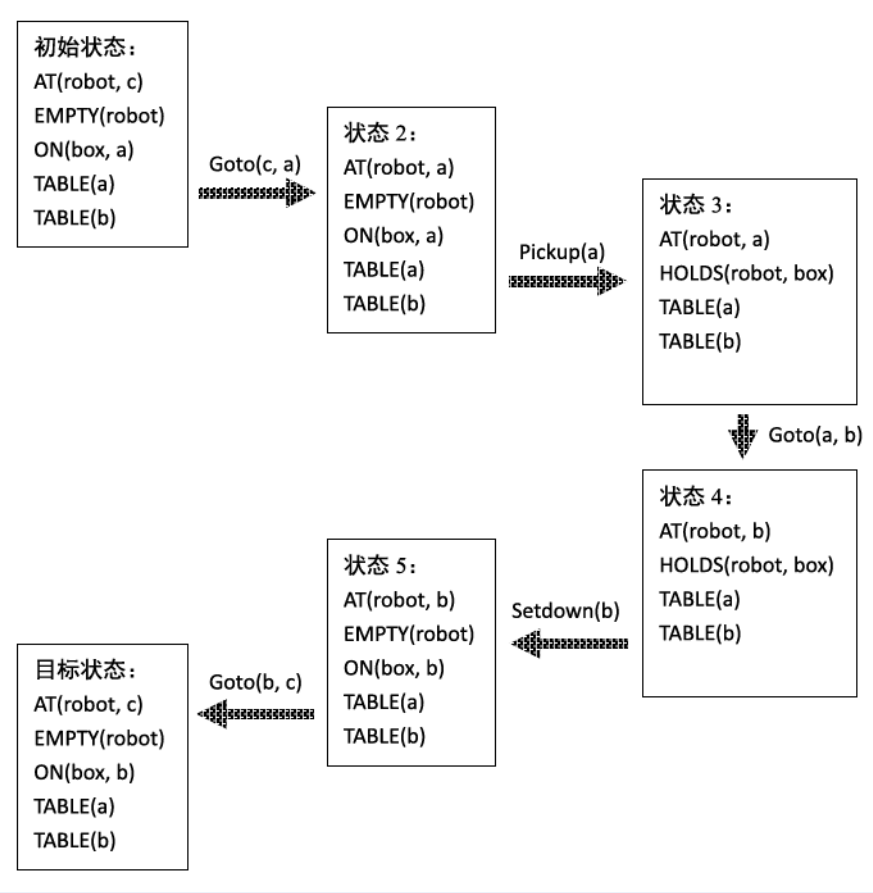


图 2.3.5　用一阶谓词逻辑表示法描述机器人移盒子的过程

◆ **用一阶谓词逻辑表示法描述猴子摘香蕉问题的过程，填写表2.3.3。**

表 2.3.3　用一阶谓词逻辑表示法描述猴子摘香蕉问题的活动记录表

项目	描述	说明
描述状态的谓词设计（包括个体域说明）		
初始状态和目标状态		

续表

项目	描述	说明
操作谓词设计		
过程描述		

2.4 搜索技术

通过学习 2.2 节我们知道，问题求解是人工智能的基本技术之一，任何复杂的问题求解不仅离不开适当的知识表示方法，而且离不开搜索技术。这里的搜索指的是广义的搜索，即从条件到答案存在着一条或多条路径，找到路径就意味着问题获得了解答，而找到路径的方法即搜索。

本节将围绕“探路”这个有趣的问题展开学习，通过“多路递归探路”和“启发式探路”应用案例，帮助同学们理解搜索的基本概念和应用方法，进一步认识人工智能中知识表示方法的应用环境。

学习目标

★ 理解多路递归及使用多路递归实现搜索的原理。

★ 了解启发式搜索及其在探路中的应用。

任务　通过解决探路问题学习搜索技术

※ 活动1　多路递归探路

在信息技术必修课中，我们已经接触了递归的概念，这里，我们将进一步研究递归，并把这一编程技术应用到探路问题上。

某函数在内存中的执行过程如图2.4.1所示。

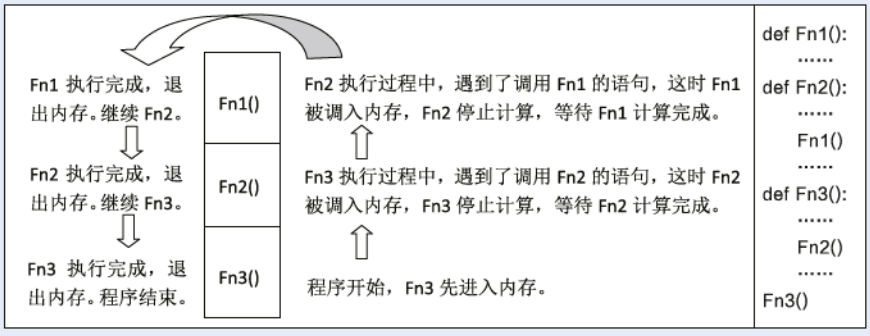


图 2.4.1　某函数在内存中的执行过程

递归调用和上述函数执行过程方式类似，只不过是函数“自己调用自己”，递归函数在定义上要求：（1）必须有参数；（2）参数必须是按照某种顺序（从大到小或从小到大）变化的；（3）函数体中必须有函数终止条件语句。

用递归函数编程求n！的代码如下所示。

```
def fact（n）：
  if n == 1:
    return 1
  return n*fact（n-1）
```

如图2.4.2所示，其递归原理是：递归 = 递推+递归终止条件+回归。

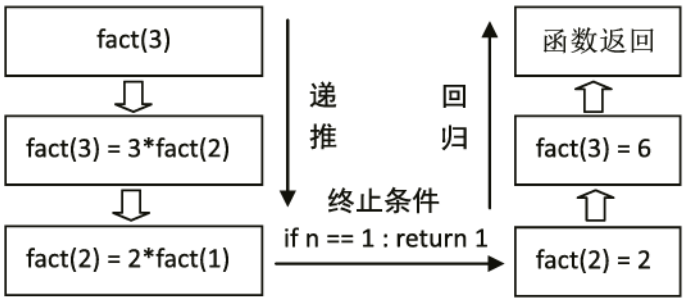


图 2.4.2　递归原理解析

在探路问题中，我们采用了多路递归的方式，也就是说，在一个递归函数中，函数多次调用了自己，当然，每次调用给出的参数是不同的。同学们可以在教科书配套资源中找到“多路递归探路”程序的完整代码，程序中使用了Python的turtle绘图包，可以形象地模拟迷宫的通路和障碍，其实，不使用turtle绘图包也可以表示迷宫中的通路。

在程序中构造的迷宫的数据结构如图2.4.3所示，它可以通过一个随机产生的二维数组表示，其中1表示墙，0表示路。

```
maze = [[1, 0, 1, 1, 1, 1, 1, 1, 1, 1],
        [1, 0, 0, 0, 1, 1, 0, 1, 0, 1],
        [1, 0, 1, 0, 0, 0, 1, 0, 0, 1],
        [1, 0, 0, 0, 1, 0, 0, 1, 0, 1],
        [1, 1, 0, 1, 0, 1, 0, 0, 1, 1],
        [1, 0, 0, 0, 1, 0, 0, 1, 0, 1],
        [1, 1, 1, 1, 1, 0, 0, 0, 1, 1],
        [1, 0, 1, 0, 0, 0, 1, 0, 0, 0],
        [1, 1, 0, 0, 1, 1, 0, 0, 1, 1],
        [1, 1, 1, 1, 1, 1, 1, 1, 1, 1],
       ]
```

图 2.4.3　迷宫的数据结构

用Python中的“小海龟”（turtle）按照上述数据结构绘制迷宫，结果如图2.4.4所示。

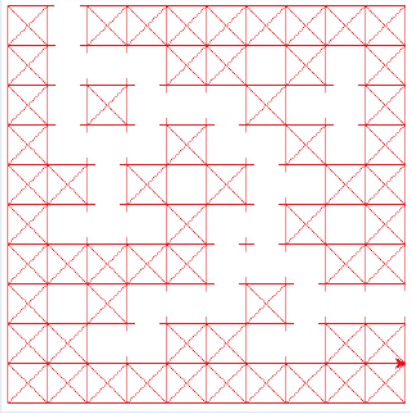


图2.4.4 迷宫的可视化设计

在迷宫中盲目搜索通路的过程可以用如图2.4.5所示的代码实现。

```
sX = 0;sY = 1;tX = 7;tY = 9
success = 0          #success是0表示寻路不成功，是1表示成功
def visit(i,j):
    maze[i][j] = 2   #对走过的路径做标记
    global success   #用全局变量记录当前是否成功
    if i==tX and j==tY:success = 1   #成功递归结束的条件
    if success != 1 and i > 0 and maze[i - 1][j] == 0:
        visit(i - 1,j)   #向上递归
    if success != 1 and i < len(maze[i])-1 and maze[i + 1][j] == 0:
        visit(i + 1,j)   #向下递归
    if success != 1 and j > 0 and maze[i][j - 1] == 0:
        visit(i,j - 1)   #向左递归
    if success != 1 and j < len(maze[i]) and maze[i][j + 1] == 0:
        visit(i,j + 1)   #向右递归
    if success != 1: maze[i][j] = 3 #不成功递归结束的条件
    return success
```

图2.4.5 迷宫的代码设计

visit函数的参数表示搜索到的位置（i、j表示矩阵的行和列），sX和sY表示迷宫入口的行列值，tX和tY表示迷宫出口的行列值，4个if语句是向四个方向递归搜索。

从代码中可以看出，函数通过多路递归调用，不断修改矩阵的值。首先是将走过的路径标记为2，如果随着递归的深入发现道路不通，则将不通的“岔路”都标记为3。可以想见，观察执行过visit函数的矩阵，里面应该有四类数值：0，表示没有试走的通路；1，表示墙；2，表示试走成功的通路；3，表示试走失败的“岔路”。其结果如图2.4.6所示，红线勾出的就是迷宫的通路。

```
in:maze[0][1],out:maze[7][9]
Show access:
[1, 2, 1, 1, 1, 1, 1, 1, 1, 1]
[1, 2, 3, 3, 1, 1, 0, 1, 0, 1]
[1, 2, 1, 2, 2, 2, 1, 0, 0, 1]
[1, 2, 2, 2, 1, 2, 2, 1, 0, 1]
[1, 1, 3, 1, 0, 1, 2, 0, 1, 1]
[1, 3, 3, 3, 1, 3, 2, 1, 0, 1]
[1, 1, 1, 1, 1, 3, 2, 2, 1, 1]
[1, 0, 1, 3, 3, 3, 1, 2, 2, 2]
[1, 1, 3, 3, 1, 1, 3, 3, 1, 1]
[1, 1, 1, 1, 1, 1, 1, 1, 1, 1]
```

图2.4.6 多路递归探路程序的运行结果

仔细阅读“多路递归探路”程序代码并填写表2.4.1。

表 2.4.1　“多路递归探路”活动记录表

序号	问题	结果记录
1	“多路递归探路”程序中四条递归调用的条件各是什么？条件的设置起到什么作用	
2	“多路递归探路”程序中，探索失败的路径被标记为3，成功探索过的路径被标记为2，程序是如何实现这样的标记方式的？尤其是在2、3相邻的情况下，程序如何区分“岔路”和“通路”	

※ 活动2　启发式探路

图2.4.7描述了猎人打猎的情景，猎人和猎物之间存在障碍物，猎人需要找到一条可以走到猎物所在地的通道。为了简化问题，我们将场景“方格化”：将有效区域划分为7*6的方格。启发式探路过程我们采用经典的A*算法。

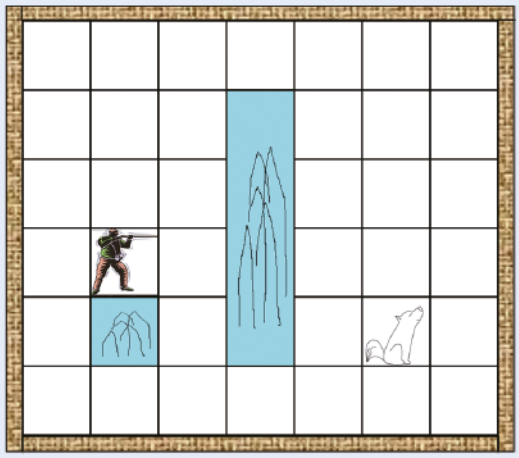


图 2.4.7　启发式探路的应用场景

A*算法需要两个表，一个表记录所有被考虑来寻找最短路径的方块（称为open表），另一个表记录不会再被考虑的方块（称为closed表）。首先在closed列表中添加当前位置，我们把这个开始点称为点“A”，然后把所有与当前位置相邻的可通行小方块添加到open列表中。现在猎人需要判断在这些选项中，哪项才是最短路径。在A*探路算法中，通过给每一个方块一个估价函数值来判断要走的方向，该值被称为路径增量。

A*算法中，估价函数F=G+H。G是从开始方块A（初始方块，即猎人开始所在的位置）到当前方块的移动量，所以从开始方块A到相邻方块的移动量为1，该值会随着与开始点距离的增大而增大。H是从当前方块到目标方块（我们把它称为点B，代表猎物）的移动量估算

值。这个常被称为探视，因为我们不确定移动量是多少，仅仅是一个估算值。为了计算出G的值，我们需要从它的前继（上一个方块）获取，然后加1，所以每个方块的G值代表了从点A到该方块所形成路径的总移动量。H值是从当前方块到终点的移动量估算值，为了简单起见，我们只计算当前方块距离方块B剩下的水平和垂直的方块数量。采用如图2.4.8所示的方法描述方块的F、G和H值。

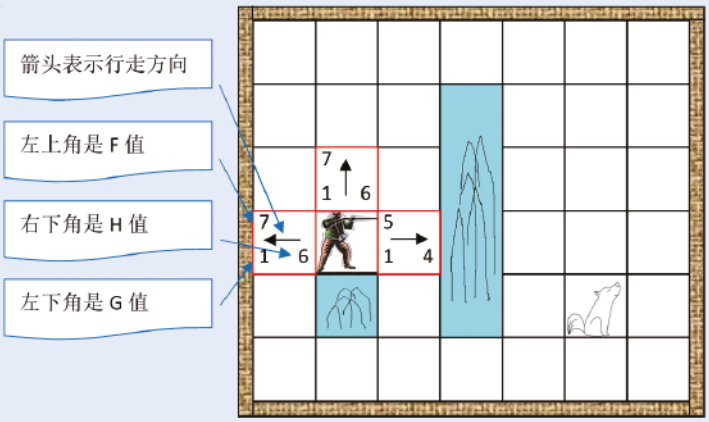


图2.4.8 F、G和H值的表示

A*算法描述

第一步：猎人确定相对于开始位置的相邻方块，如果相邻方块中包含了目标方块B，则算法结束，否则计算出它们的F值，然后把它们添加到open表中。

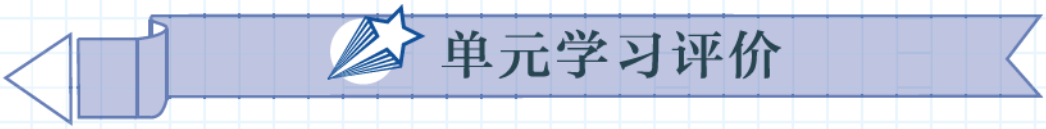
第二步：猎人选择F值最小的方块，将它从open表中删除，然后把它添加到closed表中，接着检索它的相邻方块的相关数值。

第三步：重复第一步和第二步。

同学们可以在教科书配套资源中找到“启发式探路”的完整程序，请阅读程序并填写表2.4.2。

表2.4.2 “启发式探路”活动记录表

序号	主题	结果记录	
1	按照A*算法，为猎人找出一条“最佳路径”，并且计算每走一步的F、G、H值		
2	“启发式探路”中的“启发”是什么意思		
3	启发式探路和多路递归探路相比，哪种方法更好？从给出的两个角度进行比较	智能性	
		程序实现的难易度	



单元学习评价

本单元我们学习了知识表示发展史、知识的状态—过程表示方法、知识的逻辑表示方法、搜索技术等内容，接触到了一阶谓词逻辑、产生式、语义网络、状态空间、与/或树、知识图谱、事理图谱等7种知识表示方法，还学习了与知识表示密切相关的搜索技术、专家系统等内容。

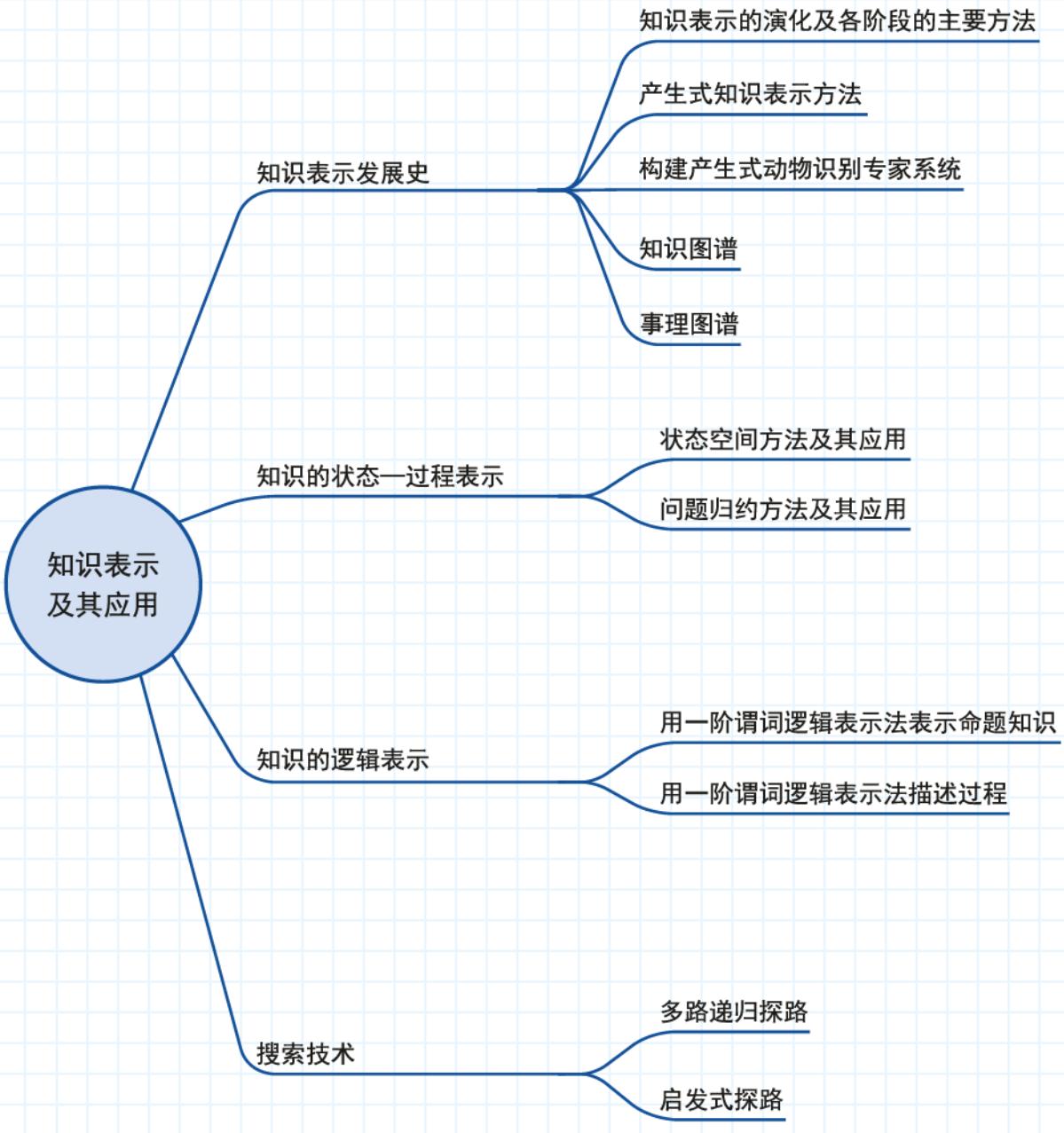
请结合本单元的学习内容，回答以下问题。

1. 知识图谱和事理图谱是当前流行的知识表示方式，请谈谈它们的相同之处和不同之处。

2. 本单元我们使用状态空间方法解决了二阶汉诺塔问题，并用状态空间图绘制了问题解的路径。请用状态空间法解决三阶汉诺塔问题，最后通过状态空间图表示它的解的路径。

3. 本单元我们使用问题归约方法解决了三阶汉诺塔问题，并用与/或树绘制了问题的解树。请用问题归约方法解决二阶汉诺塔问题，最后用与/或树绘制它的解树。

单元学习总结


知识表示及其应用
知识表示发展史
知识表示的演化及各阶段的主要方法
产生式知识表示方法
构建产生式动物识别专家系统
知识图谱
事理图谱
知识的状态—过程表示
状态空间方法及其应用
问题归约方法及其应用
知识的逻辑表示
用一阶谓词逻辑表示法表示命题知识
用一阶谓词逻辑表示法描述过程
搜索技术
多路递归探路
启发式探路

第 3 单元　机器学习与深度学习

学习能力是人类智能的根本特征，人类通过学习来提高自己的能力。学习的基本机制是设法把在一种情况下成功的表现行为转移到另一类似的新情况中去，任何具有智能的系统都必须具备学习的能力。那么机器到底是如何学习的呢？它和人类学习有什么本质区别吗？

通过对本单元的学习，我们可以知道上述问题的答案，进一步了解人工智能的相关应用及其实现方法，了解机器学习、神经网络、深度学习等基本概念及其主要应用。

本单元中，我们将结合“了解人工神经网络与机器学习”项目，围绕人工智能关键技术应用案例开展学习。

为了完成该项目，需要考虑以下问题：什么是机器学习？它与人类学习有什么区别？机器学习的主要应用领域有哪些？什么是人工神经网络？什么是深度学习以及它对人工智能的发展起到了哪些推动作用？为此，我们需要完成以下任务：

- 体验机器学习，了解机器学习的概念及其发展历程
- 了解机器学习的原理及应用领域
- 了解人工神经网络的基本原理，体验简单人工神经网络的学习过程及深度学习的应用

3.1 机器学习的起源与发展

由于机器学习的研究有助于发现人类学习的机理和揭示人脑的奥秘，故在人工智能发展早期，对机器学习的研究就处于重要的地位。本节将从一个简单的在线游戏开始，体验机器学习的特点，并通过数字化学习了解机器学习的概念、起源与发展历程。

学习目标

★ 通过在线小游戏，体验机器学习的特点。

★ 了解机器学习的概念与发展历程。

对于机器学习，同学们一定有不少疑惑：机器如何进行学习？机器学习有什么用途？机器学习和人工智能之间是什么关系？机器通过学习就能超过人类吗？为了让大家对机器学习有个感性的认识，我们先通过一个小游戏来体验和了解机器学习与人类学习的差距及机器学习的特殊性。

任务 体验机器学习，了解机器学习的概念及其发展历程

※ 活动1 通过FlappyBird小游戏体验机器学习

FlappyBird 游戏于 2013 年 5 月发布，曾经成为 iTunes 最受欢迎的免费应用软件。该游戏操作简单，只需要用一根手指来操控，点击触摸屏，小鸟就会往上飞，不断点击小鸟就会不断地往高处飞；停止点击，小鸟则会快速下降。所以若要使小鸟一直向前飞行，就要帮它躲避途中高低不平的障碍物，因为小鸟一旦碰到障碍物，游戏就会结束。如果在 PC 机上操作，当小鸟开始飞行时，单击游戏窗口的空白处，每单击一次，

小鸟就向上飞行，单击频率越高，小鸟就飞得越高。可以通过单击调整小鸟的飞行位置，让它恰好通过障碍物间的空隙。每当小鸟飞过一组障碍物时，玩家就会获得一个奖杯。其游戏界面如图 3.1.1 所示。

图 3.1.1　FlappyBird游戏界面（无监督学习模式）

分组竞赛：

1. 先从资源库中将“FlappyBird”文件夹中的三个程序文件下载到本地，分别分发给事先按班级人数分好的三个小组。

2. 第一组同学分配的程序为人工学习组（FlappyBird_ 预设版_.bcm），第二组为有监督学习组（FlappyBird 有监督学习模式 .bcm），第三组为无监督学习组（FlappyBird 无监督学习模式 .bcm）。各组准备好后，通过编程猫网站打开相应程序进行游戏竞赛。在三组游戏中，第一组游戏没有机器学习的参与，考验的是人的协调能力与判断能力，需要不停地、恰到好处地单击鼠标让小鸟通过障碍物。第二组是有监督学习模式，有机器学习的参与，运用的是 BP 反向传播算法。需要人为地单击鼠标，让小鸟通过障碍物，如果成功通过了，程序就记录下一个正反馈；如果碰到障碍物了，程序就记录下一个负反馈。通过一段时间的训练之后，AI 程序就能自主顺利地通过障碍物了。第三组游戏运用的是无监督学习模式，采用的是机器学习算法中的遗传算法，训练的时候不需要人工参与，机器自动产生一系列小鸟，从不同的位置尝试通过障碍物，如果失败，则产生下一代小鸟，重新尝试，直到通过障碍物。完成训练后，人工智能程序就可以控制小鸟自主顺利地通过障碍物，不再需要人为干预。

竞赛规则：首先，每组分别用 5 分钟熟悉游戏的玩法，然后开始

训练，时间为 10 分钟。正式开始竞赛时间为 10 分钟，由老师统一喊口令“开始”与“停止”。竞赛结束后，填写表 3.1.1，看哪一组取得了胜利。

表 3.1.1 FlappyBird小游戏竞赛得分表

组别	人数	奖杯平均数	名次
FlappyBird_人工学习组			
FlappyBird_有监督学习组			
FlappyBird_无监督学习组			

思考：在这个游戏中，人类有可能取得胜利吗？

机器学习

机器学习（Machine Learning，ML）是人工智能的一门分支科学，该领域专门研究计算机怎样模拟或实现人类的学习行为，以获取新的知识或技能，重新组织已有的知识结构，使之不断改善自身的性能。从实践的角度看，机器学习是利用过去的数据对模型进行训练，然后使用模型对将来进行预测的一种方法。

所谓模型，可以简单理解为函数。确定模型就是选择符合数据特征的函数；训练模型就是用已有的数据，通过各种优化算法确定函数的参数；参数确定后，再把新的数据代入函数求值就是使用模型的过程。

机器学习涉及概率论、统计学、算法复杂度理论等多门学科。它从学习方式上可分为有监督学习和无监督学习。有监督学习是指通过已有的训练样本（已知数据以及其对应的输出）来训练，建立起模型，再利用这个模型计算出测试数据（未知数据）的输出结果，从而实现数据的分类与判断。无监督学习事先没有训练数据样本，直接对数据进行分类建模，即没有经验和训练数据样本做参考，需要计算机自己根据整体数据特点进行分类等建模。

机器学习是人工智能的核心技术之一，是使计算机具有智能的重要途径，其应用遍及人工智能的各个领域。机器学习也是所有语音助手产品（包括Apple的Siri与Google的Now等）能够跟人交互的关键技术。除了自然语言理解，机器学习还广泛应用于图像识别、数据挖掘、专家系统、机器人和机器博弈等领域。

※ 活动2　通过数字化学习了解机器学习的发展历程

通过活动1中的“FlappyBird”小游戏，我们初步知道了什么是机器学习，那么机器学习技术的发展历程是怎样的呢？它在人工智能发展过程中的地位如何？为了解答这些问题，可以就机器学习的阶段划分、每个阶段的发展状况及标志性事件进行调查，并填写表3.1.2。

表 3.1.2　机器学习的发展历程及其标志性事件

阶段	时间	机器学习的发展情况	标志性事件
第一阶段	20世纪50年代到60年代	这一阶段属于发展的热衷时期，研究目标是自组织系统与自适应系统	诞生了模式识别新学科，形成两种重要的机器学习方法：判别函数法和进化学习
第二阶段	20世纪60年代到70年代		
第三阶段			
第四阶段			

思考：在机器学习进入第四阶段即最新阶段至今，同学们所了解的最新的机器学习应用有哪些？

在老师的指导下，设计出高效的搜索关键词，如遗传算法、强化学习等，通过在线搜索获取国内或国际机器学习学术研讨会上的相关信息，以此开展数字化学习，更好地了解机器学习的最新知识。然后分组讨论，以小组为单位在班级中进行分享和交流，看哪一组得到结论的速度最快，获得的信息最全面、最权威。

强化学习

强化学习（Reinforcement Learning，RL），又称增强学习或激励学习，在强化学习中，学习系统根据从环境中获得的反馈信号的状态（奖励/惩罚），调整系统的参数，以达到其目标的最佳动作。在计算机领域，第一个强化学习问题是利用奖惩手段学习迷宫策略。最简单的强化学习采用的方法是学习自动机（Learning Automata）。强化学习作为一种机器学习的方法，在实际中取得了很多应用，例如博弈、机器人控制等。其中，在互联网信息搜索中，也经常采用强化学习来解决搜索引擎

自动适应用户要求等问题。尽管强化学习存在很多优点，但也有不少问题，比如大多数强化学习模型针对的是单目标学习问题的决策策略，难以适应多目标多策略的学习要求。

学习自动机

在强化学习方法中，学习自动机是最普通的方法。学习自动机系统的学习机制包括自动机和环境两个模块。学习过程是根据环境产生的刺激开始的。自动机根据所接收到的刺激对环境做出反应，环境接收到该反应对其做出评估，并向自动机提供新的刺激。学习系统根据自动机上次的反应和当前的输入自动调整自身的参数。学习自动机的学习模式如图3.1.2所示。这里的延时模块用于保证上次的反应和当前的刺激同时进入学习系统。

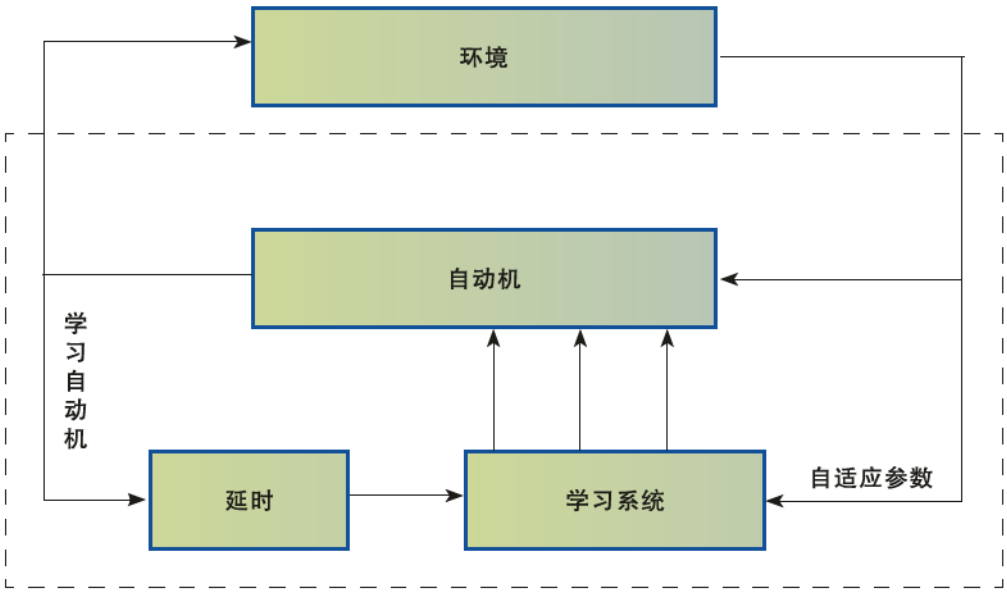


图 3.1.2　学习自动机的学习模式示意图

3.2　机器学习的原理、分类与内涵

机器学习是人工智能最为重要的研究领域之一。机器学习的应用包括语音识别、图像识别、数据挖掘、专家系统、智能机器人等。一个系统是否具有学习功能，已成为判断它是否具有“智能”的重要标志。本节将围绕机器学习的应用体验展开学习，通过“感知器的运行原理”“借助决策树设计‘等人’决策”等应用案例，帮助同学们了解机器学习的基本原理、分类与内涵，知道人工智能、机器学习与深度学习之间的关系，体验机器学习在人工智能领域的应用。

学习目标

★ 了解机器学习的应用领域。

★ 通过感知器与“等人”决策的运行过程了解机器学习的原理。

下面我们先从感知器的运行与“等人”决策的制订开始了解机器学习。

任务　了解机器学习原理与应用领域

※ 活动1　以感知器为例，了解机器是怎么学习的

本活动以多因素判断小明是否会去看电影的决策过程为例，引导同学们了解感知器的实现原理。

● 感知器

感知器是美国计算机科学家罗森布拉特（Frank Rosenblatt）于1957年提出的。

单层感知器是一个具有一层神经元，采用阈值（Threshold）激活函数的前向网络。通过对网络权值的训练，可以使感知器对一组输入矢量的响应输出为 0 或 1 ，从而实现对输入矢量分类的目的。图3.2.1显示的是单个神经元模型。

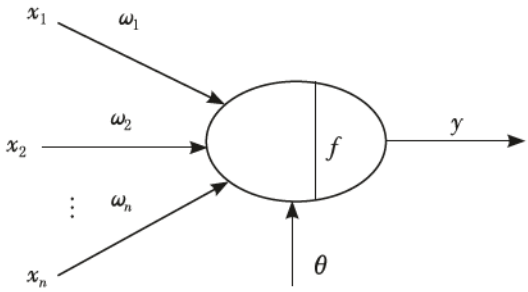


图 3.2.1　单个神经元模型

其中，x_1，x_2，…，x_n为输入信号，ω_1，ω_2，…，ω_n为各输入变量的权重，θ 为阈值，f 是激活函数。权重的取值代表相应输入变量对于输出结果的重要性。神经元的输出结果（0或者1）由分配权重后的所有变量值和其对应的权重的乘积的总和与阈值相比之后决定。权重和阈值都是实数，可以用下列代数形式精确地表示一个神经元的输出结果和各参数之间的关系。

$$y=\begin{cases}0 & \sum_{i=1}^{n}\omega_i x_i<\theta\\ 1 & \sum_{i=1}^{n}\omega_i x_i\geqslant\theta\end{cases}$$

实际输出还要通过激活函数进行转换，所以最终输出结果可以表示为：

$$y=f\left(\sum_{i=1}^{n}\omega_i x_i-\theta\right)$$

通过神经元模型及相关公式，我们对感知器的运行原理已经有了初步印象。感知器能较容易地实现逻辑与、或、非运算，两层感知器就能解决异或问题，完成比较复杂的逻辑运算。

通常采用sigmoid函数作为激活函数。sigmoid函数是一个在生物学中常见的S形曲线函数，也称为S形生长曲线，它把可能在较大范围内变化的输入值挤压到（0，1）输出值范围内，是一个良好的阈值函数。sigmoid函数的表达式为 $\text{sigmoid}(x)=\frac{1}{1+e^{-x}}$。

小明喜欢的电影周末将要上映，此时他正在犹豫是否去看电影。以下因素将或大或小地影响小明的决定：

（1）周末是否会加班？

（2）朋友会不会陪他去？

（3）是否能选到好的座位？

小明分别用 x_1、x_2 和 x_3 来表示这三个影响因素，并把它们作为感知器的三个输入变量，然后用布尔值True（1）和False（0）来表示是否去看电影，作为感知器的输出值。如果周末不加班，则令$x_1=1$；如果加班，则令$x_1=0$。类似地，如果朋友陪他去，则令$x_2=1$，否则 $x_2=0$。假设小明非常想看这场电影，只要周末不加班，他就想去

看电影，几乎不太想考虑其他因素，这时我们可以借助感知器帮小明做决定，将是否加班的权重值设置得高点，如设$\omega_1 = 0.6$，$\omega_2 = 0.2$，$\omega_3 = 0.2$（假设各因素权重之和为1）。ω_1被赋予的值越大，表示加班与否对小明看不看电影影响越大，比朋友是否陪他和是否能选到好座位重要得多。

如果在这个模型中，我们将感知器的阈值设为0.5。那么，根据上述函数关系，$x_1*\omega_1=0.6$，结果已经大于阈值0.5，即不加班就输出1，加班则输出0。朋友是否要去，以及是否有好的座位已经影响不了结果了。而如果小明要加班，即便有朋友陪，座位也很好，根据上述公式$x_1*\omega_1+x_2*\omega_2+x_3*\omega_3=0*0.6+1*0.2+1*0.2=0.4$，结果小于阈值0.5，感知器输出0，小明不会去看电影。

通过改变权重和阈值，我们可以得到不同的决策结果。现在，小明想把阈值改为0.7，那么请同学们帮小明想想，此时感知器会做出什么反应？请结合上述计算公式，帮小明做出判断，并尝试通过多次修改阈值来更改输出结果，完成表3.2.1。

表 3.2.1　感知器阈值与相应的行为

序号	阈值	满足什么条件小明会去看电影
1	0.5	只要不加班就去看电影
2	0.7	
3	0.2	

※ 活动2　借助决策树设计“等人”决策

相信大家都有等人的经历。小王有个朋友叫小明，小明经常在有约时迟到。有一次小王与小明约好下午3 点在麦当劳见面，刚要出门时小王想到一个问题：我现在出发合适吗？我会不会到了约定地点后，又要花上30 分钟等他？于是小王决定采取一个策略来解决这个问题。

小王把以往跟小明相约的经历在脑海中回忆了一遍，看看跟小明相约的次数中，小明迟到的次数占了多大比例，利用这个比例来预测他这次迟到的可能性。如果这个比例超出了小王心里所能接受的阈值，那就断定他这次又会迟到。例如，假设小王和小明约过五次，小明迟到的次数是一次，那么他按时到的比例为80%，如果小王心中的阈值为70%，那小王就会认为这次小明迟到的可能性比较小，因此就会按时出门。如果小明在五次相约中迟到的次数占了四次，也就是他按时到达的比例为20%，由于这个值低于小王能接受的阈值，因此小

王会选择推迟出门的时间。这个方法可称为经验法，即利用以往所有相约的数据来判断本次相约是否会迟到，因此也可以称为依据数据做出判断。事实上，机器学习的本质就是依据以往的数据对结果做出预测。

刚才小王只考虑了“频次”这一种属性，“频次”就是一个自变量，在真实的机器学习模型中，至少考虑一个自变量和一个因变量。因变量是我们希望预测的结果，在这个例子里就是小明迟到与否的判断。自变量也就是用来预测小明是否迟到的量。假设小王把时间也作为自变量，这个模型中就有了两个自变量，比如小王发现小明所有迟到的日子基本都是星期五，而在非星期五情况下他基本不迟到，于是小王可以建立一个模型，来模拟小明迟到与否和日子是否是星期五的概率。图3.2.2所示是一个决策树模型，也是最简单的机器学习模型。由于这种决策分支画成图形很像一棵树的枝干，故称为决策树。

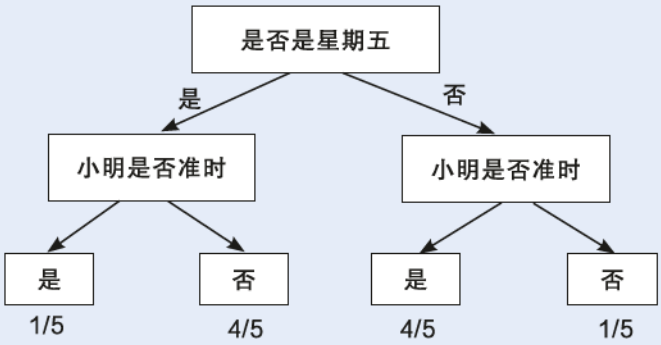


图 3.2.2　决策树模型的例子

自变量的个数还可以继续增加。例如小明迟到的部分原因是他开车时经常遇到堵车的情况，那么就要综合考虑这些信息，建立一个更复杂的模型，这个模型在原来的基础上增加了一个自变量。现在，我们考虑得再复杂一些，如果小明的迟到跟天气也有一定的关系，如下雨时更容易迟到，这时就需要考虑更多自变量。请同学们模仿图3.2.2，将上述情况用决策树模型在表3.2.2中画出来。

表 3.2.2　多变量决策树模型的练习

主要自变量： 是否是星期五、是否开车	主要自变量： 是否是星期五、是否开车、是否天气不好

思考：有多个自变量的情况下，自变量在决策树中出现的先后次序对结果是否有影响？

如果上述自变量都有具体的数值，并将小明每次迟到的时间跟

雨量的大小以及前面考虑的自变量统一建立一个模型，就可以比较准确地预测小明迟到的具体时间，知道他大概会迟到几分钟。但这种情况下，采用决策树进行分析的方法就不太适当了，因为决策树只能预测离散值（离散值指一批分散、不连续的数据集）。针对这种情况，可以采用机器学习中的另一种常用方法即线性回归法建立预测模型。首先获取某个时间段小明出门的全部数据，将所有的数据按对应自变量和因变量输入计算机，由计算机生成回归预测模型，然后根据生成的预测模型，录入当前各自的变量值，预测小明本次出门时间。事实上，计算机执行这些辅助决策的过程就是机器学习的过程。

回归分析（Regression Analysis）是确定两种或两种以上变量间相互依赖的定量关系的一种统计分析方法。按照自变量和因变量之间的关系类型，可分为线性回归和逻辑回归。线性回归指因变量和自变量之间的关系是线性相关的，即近似直线关系。按照自变量的多少，可分为一元线性回归和多元线性回归。一元线性回归只有一个自变量，是最简单的线性回归分析法；若包括两个或两个以上自变量，则称为多元线性回归。

决策树

决策（Decision）是根据信息和评价准则，用科学方法寻找或选取最优处理方案的过程或技术。对于每个决策，都有可能引出两个或多个事件，导致不同的结果或结论。将这种决策分支用一棵搜索树表示，即决策树。

决策树（Decision Tree）是在已知各种情况发生概率的基础上，通过构造树状结构来判断其结果可行性的一种分析方法，是一种比较直观的概率分析图解法。决策树是一个预测模型，是机器学习中一种十分常用的分类方法，它代表的是对象属性与对象值之间的一种映射关系。

机器学习的分类

学习是一项复杂的智能活动，学习过程与推理过程是紧密相连的，按照学习中对推理的使用程度，可将机器学习所采用的策略分为机械学习、示教学习、类比学习和示例学习四种。机械学习是最简单的学习策略，不需要任何推理，主要通过系统实现存储、检索与记忆；示教学习在接受外部知识时需要少量的推理、翻译和转化工作，比如专家系统就属于此类学习方法；类比学习系统只能得到与所完成任务类似的有关知识，因此，它比上述两种学习策略需要更多的推理；示例学习则需要对所接受的外部知识进行分析、总结和推广，以此得到完成任务的一般性规律，并在进一步的工作中验证或修改这些规律，因此需要用到更多的推理。一般而言，学习中所用到的推理越多，系统的智能水平就越高。

根据不同的学习策略，大致可将机器学习分为归纳学习、类比学习、强化学习、决策树学习、统计学习、神经网络学习等。

归纳学习

归纳（induction）是人类拓展认识能力的重要方法，是一种从个别到一般、从部分到整体的推理行为。归纳推理是应用归纳方法，从足够多的事例中归纳出一般性知识，提取事物的一般规律，即从个别到一般

的推理。归纳学习（Induction Learning）是应用归纳推理进行学习的一种方法，包括有监督学习和无监督学习。

类比学习

类比（Analogy）是一种很有用和很有效的推理方法，它能清晰、简洁地描述对象间的相似性，也是人类认识世界的一种重要方法。类比推理是由新情况与已知情况在某些方面的相似性来推出它们在其他方面的相似性。类比学习（Learning By Anology）就是通过类比推理，即通过对相似事物加以比较所进行的一种学习。

类比学习的过程大致可分为四个阶段：（1）输入一组已知条件和一组未完全确定的条件（新问题）；（2）对输入的两组条件，按某种相似性的定义寻找两者可类比的对应关系；（3）根据相似变换的方法，将已有问题的概念、特征、方法、关系等映射到新问题上，以获得待求解新问题所需的新知识；（4）对类比推理得到的新问题的知识进行校验，验证正确的知识存入知识库中，而暂时还无法验证的知识则作为参考性知识置于数据库中备用。

统计学习

统计学习（Statistical Learning）是指基于数据构建概率统计模型并运用模型对数据进行预测与分析。统计学习方法的三要素包括模型的假设空间、模型选择的准则以及模型学习的算法。统计学习的方法比较多，如逻辑回归、支持向量机等。

※ 活动3　了解机器学习的应用领域

通过前面的学习，我们了解了各种不同的机器学习方法，它们各自都有不同的应用领域，请同学们以“机器学习的应用领域”为主题上网检索，找出那些与机器学习相关的应用领域及其典型应用，将其填入表3.2.3中。

表3.2.3　机器学习的应用领域与典型应用

序号	应用领域	典型应用
1	计算机视觉	人脸识别、文字识别等
2		
3		
4		

思考：机器学习与人类学习是否有共同之处？

机器学习与人类学习的对比

人类在成长、生活过程中积累了很多经验，定期对这些经验进行“归纳”，便获得了生活的“规律”。当人类遇到未知的问题或者需要对未来进行“预测”的时候，可以使用这些“规律”，对未知问题与未来进行“预测”，从而指导自己的生活和工作。

机器学习就是对人类学习的模拟，机器学习中计算机处理数据、建立模型的过程相当于人类学习、归纳经验的过程，然后应用这些模型去解决很多复杂的实际问题。从整体上说，机器学习中的“训练”与“预测”过程可以对应到人类的“归纳”和“推测”过程。机器学习与人类学习的对比如图3.2.3所示。

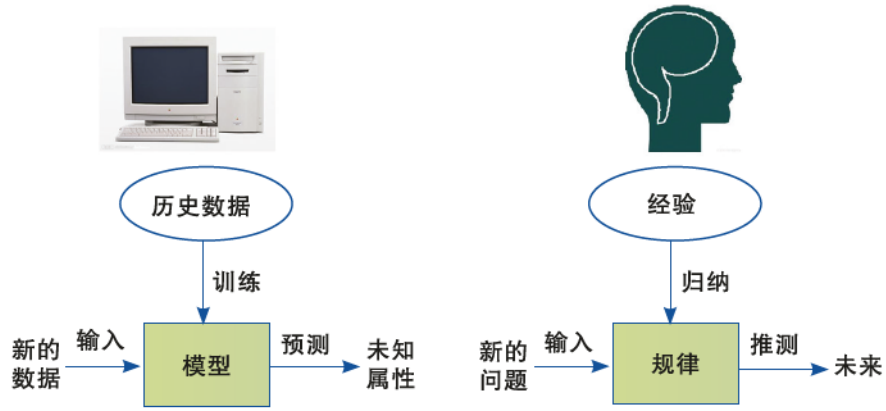


图 3.2.3　机器学习与人类学习的对比

机器学习的应用领域

机器学习的应用涉及计算机视觉、自然语言处理、模式识别、统计学习与数据挖掘等领域。

计算机视觉

计算机视觉是使用计算机及相关设备对生物视觉的一种模拟，它主要通过对采集的图片或视频进行处理以获得相应的信息。在人工智能领域，与计算机视觉相关的应用非常多，例如图像识别、人脸识别以及车牌识别等。

自然语言处理

自然语言处理是人工智能领域中的一个重要方向。它研究有助于人与计算机之间用自然语言进行有效通信的各种理论和方法。自然语言处理重点研制能有效地实现自然语言通信的计算机系统，特别是其中的软件系统。如何利用机器学习技术进行自然语言的深度理解，一直是人们关注的焦点。

模式识别

模式识别是人类的一项基本智能，在日常生活中，人们经常进行“模式识别”。人们在观察事物或现象的时候，常常要寻找它与其他事物或现象的不同之处，并根据一定的目的把各个相似的但又不完全相同的事物或现象组成一类。字符识别就是一个典型的例子。例如，数字“4”可以有各种写法，但都属于同一类别。模式识别的任务，就是模仿人类的这种识别能力。

在计算机科学与人工智能领域，模式识别是指对表征事物或现象的各种形式的（数值的、文字的和逻辑关系的）信息进行处理和分析，实现对事物或现象进行描述、辨认、分类和解释的过程。模式识别的典型应用包括文字识别、语音识别、指纹识别以及医学诊断等。

统计学习与数据挖掘

统计学习是与机器学习高度重叠的学科，因为机器学习中的大多数方法来自统计学。两者的主要区别在于，统计学习侧重统计模型的发展与优化，偏理论；而机器学习侧重解决问题，偏实践。因此，机器学习重点研究学习算法在计算机上执行的效率与准确性的提升。

数据挖掘是指从大量的数据中通过算法搜索隐藏于其中的信息的过程。数据挖掘主要利用数据库技术等存储、管理数据，并利用机器学习等方法分析数据。数据挖掘中的大部分算法都是机器学习的算法在数据库中的优化，并利用统计分析技术、人工智能技术的应用程序，把机器学习中复杂的技术封装起来，使人们不用自己掌握这些技术也能完成同样的功能，并且更专注于自己所要解决的问题，从而提高效率。

拓展练习

开放式计算机视觉库OpenCV中的人脸识别库可以用于实现人脸识别，请将自己的单人照与合照分别处理后，调用教科书配套资源中提供的人脸识别程序，看看计算机是否能通过单人照从合照中找到对应的人脸，从而体验机器学习的实际应用。

OpenCV是一个开源的跨平台计算机视觉运算程序库，可以运行在Linux、Windows等操作系统上。它由一系列C函数和C++类构成，提供了Python、MATLAB等语言的接口，实现了图像处理和计算机视觉方面的很多通用算法。

图像矩阵

图像数字化处理后可以用矩阵来表示，在计算机数字图像处理程序中，通常用二维数组来存放图像数据。比如灰度图像的像素数据就是一个矩阵，矩阵的行对应图像的高，矩阵的列对应图像的宽，矩阵元素的值就是像素的灰度值。

上述程序的实现过程如下。

（1）读取两张图像（一张为单人照，另一张为包括单人照中人物的合照），生成图像矩阵。

（2）以这两个图像矩阵为基础，调用OpenCV 的相关函数完成人脸定位。

（3）读取图像的人脸区域，生成人脸图像矩阵，并将人脸矩阵转换为灰度图。

（4）比较分析人脸图像矩阵，通过相关的计算函数找到最相近的人脸。

更详细的实现过程参见源程序的注释。

3.3 人工神经网络与深度学习

基于人工神经网络的机器学习是一类新的机器学习方法。本节我们要了解人工神经网络的基本概念，并结合实例体验人工神经网络的实现过程，以及深度学习技术的一般应用。

学习目标

★ 了解人工神经网络的基本原理。

★ 体验简单人工神经网络学习的实现过程。

★ 了解深度学习技术的一般应用。

生物体中的神经网络通常是指由大脑神经元、细胞和突触等组成的网络，神经网络的主要作用是产生意识，帮助人们进行思考和行动。那么，什么是人工神经网络？两者之间有什么区别与联系？什么是深度学习？其工作机理如何？

任务　了解人工神经网络的基本原理，体验简单人工神经网络的学习过程及深度学习的应用

※ 活动1　了解人工神经网络

按照表 3.3.1 所示的主题划分小组，在老师的指导下，设计搜索关键词，如人工神经网络，开展数字化学习，了解神经网络的相关知识，并按表所示完成相关主题的知识获取、筛选与整理，然后以小组为单位在班级中进行分享与讨论。

表 3.3.1　生物神经网络与人工神经网络的区别与联系

	生物神经网络	人工神经网络	补充说明
基本概念			
相互联系			
主要区别			
……			

人工神经网络

在机器学习和认知科学领域，人工神经网络（Artificial Neural Network，ANN）简称神经网络（Neural Network，NN）或类神经网络，是一种模仿生物神经网络（动物的中枢神经系统，特别是大脑）的结构和功能的数学模型或计算模型。神经网络由大量的人工神经元联结构成。大多数情况下，人工神经网络能在外界信息的基础上改变内部结构，是一种自适应系统。

根据计算方法的不同，目前常用的人工神经网络可分为卷积神经网络（CNN）、循环神经网络（RNN）和深度神经网络（DNN）等。

神经网络的逻辑结构

在3.2节中，我们学习了具有一层神经元的单层感知器的运行原理。事实上，感知器就是人工神经网络中的一种典型结构。神经网络的逻辑结构包括输入层、隐藏层和输出层。输入层负责接收信号，隐藏层负责对数据进行分解与处理，最后的结果被整合到输出层。每层中的一个圆代表一个处理单元，也可看成一个神经元，若干个处理单元组成一个层，若干个层组成一个网络，即神经网络。或者说，每个处理单元就是一个逻辑回归模型，逻辑回归模型接收上层的输入，把模型的预测结果作为输出传输到下一层的输入。这样，神经网络就可以完成非常复杂的非线性分类。

具有一个输入层、一个输出层、一个隐藏层的简单神经网络结构模型如图3.3.1所示。

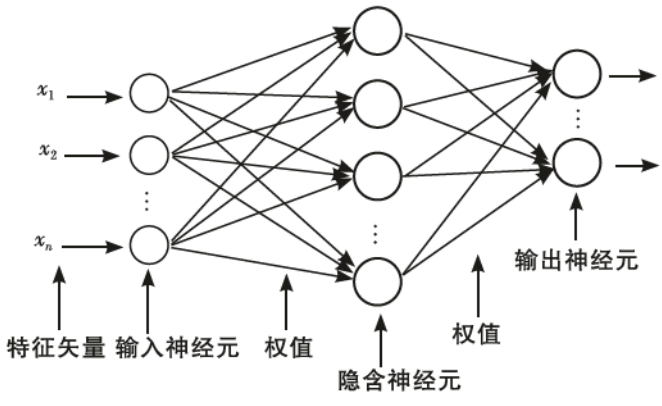


图 3.3.1　简单的神经网络逻辑结构图

神经网络的应用领域

神经网络的应用领域主要包括模式识别、信号处理、知识工程、专家系统、优化组合和机器人控制等。随着神经网络以及相关理论与技术的不断发展，神经网络的应用正逐步发展。尤其是随着多隐藏层神经网络即深度学习技术的突破，让曾经一度被冷落的神经网络应用呈现井喷式发展，这与当前人工智能的驱动因素即大数据、算法和算力的应用与提升有着密不可分的联系。

※ 活动2　体验简单人工神经网络的学习过程

我们通过对感知器知识的学习得知，变换权重和阈值，可以得到不同的决策模型，这也是感知器通过将各种影响因素进行加权求和来制订决策的过程。多层感知器组成了感知器网络，感知器网络能进行更加复杂的决策。人工神经网络就是通过学习来解决问题的人工神经元网络，与感知器类似，调整神经元的权重和阈值会对输出结果产生影响，从而通过逐步更改权重和阈值来让网络按照期望的方式发展。接下来我们通过一个简单的人工神经网络程序来体验其工作过程。

请调试和运行教科书配套资源中的简单的人工神经网络程序（network.py），通过简单的样本集让计算机进行学习（训练），然后测试并计算输出数据，从而体验人工神经网络的学习与应用过程。

这是一个由Python语言数学库NumPy基本函数构成的简单的人工神经网络程序。该程序只设置了具有三个输入（Input）和一个输出（Output）的单个神经元，相对简单。如表3.3.2所示，大家可以试着改变训练集数据、测试数据及训练迭代次数，看看会得到什么样的结果，并思考类似的人工神经网络在生活中的应用。

表3.3.2　简单人工神经网络的训练测试数据与输出结果对照表

序号	训练集数据	测试数据	训练迭代次数	输出结果	结果与预期是否相符
1	输入：[0, 0, 1], [1, 1, 1], [1, 0, 1], [0, 1, 1] 输出：[0, 1, 1, 0]	[1, 0, 0]	10000次	0.99993704	符合
2					
3					
4					

传统的计算机程序不具有学习功能，而人工神经网络却能自己

“学习”。这里实现的只是一个单一的神经元，试想一下，如果我们把数百乃至数百万的神经元连接起来，其学习效果将多么强大！

简单人工神经网络的学习过程

为了模拟人工神经网络的学习过程，这里提供了五个数据样本，前四个样本（Example 1至Example 4）作为训练集，第五个样本（New Situation）作为测试用集，如表3.3.3所示。分析该表中的数据，我们不难看出测试数据的输出结果应该是“1”，因为从前四个样本中可以发现输出结果总是和输入数据中的第一列一致。但如何通过神经网络让机器自己找到规则，判断出测试数据的输出结果呢？下面我们一起来看看这个过程是如何实现的。

表 3.3.3　一个简单人工神经网络的训练与测试样本

样本名称	Input			Output
Example 1	0	0	1	0
Example 2	1	1	1	1
Example 3	1	0	1	1
Example 4	0	1	1	0
New Situation	1	0	0	?

该人工神经网络的学习（训练）过程是：首先设置每个权重的初始值为一个随机数字，取一个训练样本的输入值，通过前述神经元与权重关系的公式计算神经元的输出，并通过sigmoid函数做归一化处理，使输出结果控制在0到1之间，称为可信度。可信度越接近1说明正确率越高，相反则越低。在此过程中，神经元会根据每次输出计算误差，即神经元的输出与训练样本中的期待输出之间的差值，不断调整对应的权重值，直到输出结果即sigmoid函数值接近0或1的时候，训练基本完成。

此例的运行结果中，输出的sigmoid函数值为0.99993704，已经非常接近1了，说明结果的正确率较高。

上述人工神经网络程序的学习（训练）和测试的结果如图3.3.2所示。

```
F:\deeplearning\simple-neural-network-master>python test.py
神经网络初始化时随机分配的权重：
[[-0.16595599]
 [ 0.44064899]
 [-0.99977125]]
训练后对应的权重：
[[ 9.67299303]
 [-0.2078435 ]
 [-4.62963669]]
预测测试数组：  [1, 0, 0] 的结果为：
[ 0.99993704]
```

图 3.3.2　测试结果输出界面

首先，神经网络程序对自己赋予随机权重，然后使用训练集反复训练并对权重做相应的调整。经过神经网络模型训练并调整后的权重分别为9.67、–0.2、–4.6。

学习过程完成后，程序读入测试数据[1, 0, 0]，计算后的输出结果为0.99993704，非常接近1，符合预期。

※ 活动3　了解深度学习

请同学们通过上网检索和在线学习，在表3.3.4中记录深度学习在计算机视觉与自然语言处理等领域的典型应用，并补充说明该领域应用的国家、机构、相关人物等信息。例如，深度学习在自然语言处理领域的典型应用有机器翻译、语音合成、语音识别、智能聊天系统等，其中科大讯飞的语音识别技术很有代表性。

表3.3.4　我所了解的深度学习及其应用

应用领域	典型应用	相关说明
自然语言处理	机器翻译、语音合成、语音识别、智能聊天系统等	例如，我国科大讯飞的语音识别技术很有代表性

深度学习

深度学习的概念由欣顿（Hinton）等人于2006年提出。深度学习的概念源于人工神经网络的研究，它的出现是人工智能发展史中的一个里程碑，可以说，深度学习携手大数据引领了人工智能发展史上的第三次热潮。深度学习的目的是建立、模拟人脑进行分析学习的神经网络，它模仿人脑的机制来解释数据。深度学习的应用领域很广，目前业界许多图像识别技术、语音识别技术的进步都源于深度学习的发展。这里所说的深度学习属于机器学习的范畴，它有别于教育学和心理学领域的“深度学习”。如图3.3.3所示，含有多个隐藏层的多层神经网络就是一种深度学习结构，深度学习通过组合低层特征，形成更加抽象的高层表示属性类别或特征，从而发现数据的分布式特征。

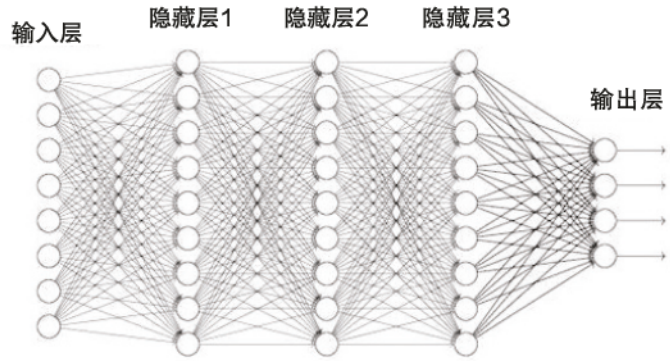


图 3.3.3　多隐藏层的神经网络结构

相对于传统的基于神经网络的学习，深度学习更注重“学习”，主要表现在：（1）强调了模型结构的深度，通常具有数层、数十层甚至更多的隐藏层节点；（2）明确突出了特征学习的重要性，也就是说，通过逐层特征变换，将样本在原空间的特征表示变换到一个新特征空间，从而使分类或预测更加容易；（3）更适用于处理大样本数据。与人工规则构造特征的方法相比，利用大数据来学习特征，所获取的信息会更加丰富。

● 深度学习框架

随着深度学习技术的应用推广，越来越多的机构推出了功能强大的深度学习框架，如谷歌的开源主流深度学习框架TensorFlow、百度开源的深度学习框架PaddlePaddle，以及Facebook发布的开源深度学习框架Caffe2等。随着时间的推移，深度学习框架在轻量、模块化和扩展性等方面正不断进步。

除了深度学习框架，还有很多通用的机器学习框架，比如Sklearn、Theano以及Spark MLlib等。其中，Theano框架作为机器学习框架的“元老”，在它之上已经建立起了很多开源的深度库，包括Keras、Lasagne和Blocks等。这些支持库的建立使得Theano功能更加强大。

TensorFlow目前已在世界各国的知名科技公司广泛应用。基于TensorFlow框架和MNIST数据集实现的手写体数字识别是深度学习技术的一种经典应用。

※ 活动4　体验TensorFlow深度学习应用——识别手写体数字

MNIST是由NIST（美国国家标准与技术研究所）收集的手写数字数据库，它有60000个训练样本集和10000个测试样本集。学生可以直接从网上下载获取MNIST数据库，MNIST库中的数字图片如图3.3.4所示。要体验识别手写体的数字，还需要安装一个基于Python的用于处

理快速线性代数的库NumPy。

简单来说，识别手写体数字的实现过程可分为以下几步。

图3.3.4 MNIST库中的数字图片

第一步，数据封装。通过TensorFlow将MNIST数据划分为训练、验证、测试三个数据集并封装成一个类（类是面向对象编程技术的一个基本概念，它将数据以及这些数据上的操作封装在一起，并可用它为模板创建具体的对象），这个类会自动下载并转化MNIST数据的格式，将原始的数据包中的数据解析成训练和测试神经网络时使用的格式。

第二步，数据训练。通过TensorFlow程序训练神经网络，此时用的是训练集数据。

第三步，判断模型效果。在模型构建过程中，采用不同的隐藏层，是否使用激活函数，以及不同的优化方式都会影响结果的正确率。在程序中可以使用验证数据集判断模型效果。

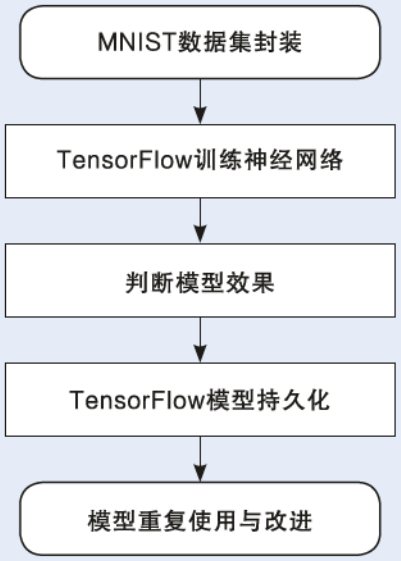


图3.3.5 基于TensorFlow的手写体数字识别实现流程图

第四步，保存训练模型。为了重复使用训练结果，可以通过TensorFlow程序将训练得到的神经网络模型持久化并保存。

基于TensorFlow的手写体数字识别实现流程如图3.3.5所示。

基于TensorFlow的手写体数字识别程序可以从教科书配套资源中下载并调试运行。该源程序主要包含两大模块。第一个模块主要是实现训练并保存模型（trainModel.py）。在该模块中，首先载入MNIST数据集，然后创建模型，通过训练达到预定目标后保存模型。第二个模块主要是加载模型，识别手写数字（recogniseNum.py）。请大家尝试多次运行程序并记录不同迭代次数对应的数字识别准确率，完成表3.3.5。运行程序时，注意当迭代次数接近多少以后，识别准确率开始趋于稳定，并记下最终的模型识别准确率。

表 3.3.5　迭代次数与识别准确率对照表

序号	迭代次数	识别准确率	结论
1	100	82%	识别准确率不稳定，变化率上升幅度较大
2	500		
3	1000		
4	1500		

在实际操作过程中会发现，人工神经网络的结构对最终结果的影响很大，使用了激活函数和隐藏层的神经网络要远远好于没有激活函数或者没有隐藏层的神经网络。为了更直观地感受这个过程，请尝试运行教科书配套资源中的手写体数字识别程序，体验深度学习的强大功能。同时，上网体验“百度识图”的应用，思考百度识图的实现过程，并上网搜索基于TensorFlow框架的其他深度学习的应用案例及其实现过程，如TensorFlow识别字母扭曲干扰型验证码的实现过程等。

深度学习技术的发展及其局限性

机器学习是人工智能的子领域，深度学习又是机器学习的子领域。从技术的发展起步上来说，也是先有人工智能，再有机器学习与深度学习。随着计算机软硬件技术以及人工智能技术的发展，人工神经网络的隐藏层已经可以从一层扩展为数十层乃至数百层，从而使基于人工神经网络的深度学习走向实用成为可能，并在许多领域都取得了有效的突破，如在视觉识别与语音识别上显著地突破了原有机器学习技术的界限，无论是“谷歌大脑”还是“百度大脑”，都是由海量级的深度学习网络构成的。

尽管深度学习在人工智能领域取得了一次又一次的突破，但深度学习技术并不是万能的。深度学习的应用离不开大数据和计算模型，如果提供给神经网络的是不准确或者不完整的数据，将会得到错误的结果，而这个错误的结果可能会让人很尴尬。如果这个错误的结果是发生在无人驾驶这样的应用场合，那就是危险的甚至是致命的。类似地，对于躲避汽车这样的行为，人类只需要被告知一次，就能从简单、少量的例子中概括出将来可能发生的事情，这是因为人类具有推理能力，并且能够想象（模拟）后果，因此不需要用生命去尝试很多次是否或如何避开汽车。虽然深度学习在大数据的学习基础上可以达到令人惊讶的结果，但由于其原理主要基

于复杂的模式识别，而非推理，对于个案的处理还存在很多问题。比如，前面同时出现一个人和一堵墙时，机器可能会出现人类永远不会出现的错误。很多科学家也认为人们在很多方面都高估了深度学习的功能。尽管如此，深度学习仍然是今后人工智能领域研究的热点，还会有更多新的深度学习技术被开发出来，基于深度学习的应用也会有革命性的进步。虽然目前深度学习的原理还停留在复杂的模式识别层面上，但它依然是很多领域最有效的技术应用，并极大地推动着人工智能的进步。

拓展练习

在前面的学习中，我们知道了最简单的神经网络实现过程，现在我们将通过一个手写体数字识别案例让大家体验基于BP算法的人工神经网络的应用。源程序可以从教科书配套资源中下载，其代码主要包括两个文件saveModel.py（创建模型）和numPredict.py（数字识别结果预测），实现过程大致可分为以下几步。

（1）创建并训练模型。

加载开源标准数据集 mnist.pkl.gz，调用相关函数并设定好迭代次数、样本数量、学习率等参数，然后开始训练神经网络。

（2）保存模型。

通过修改参数，比如迭代次数、样本数量、学习率等，使模型效果不断优化，并保存识别率最高的结果。

（3）预测结果。

调用模型对测试图片进行预测。基于BP神经网络学习的手写体数字识别流程图如图3.3.6所示。

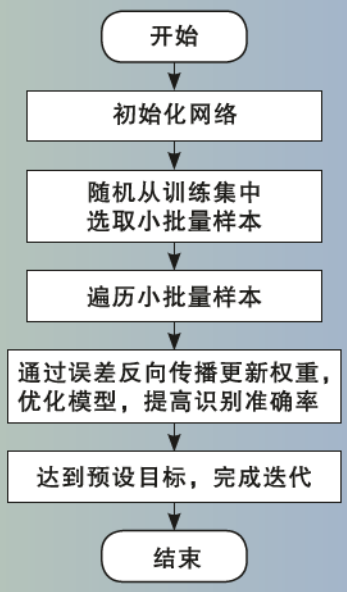


图 3.3.6 BP算法流程图

为了更好地体验BP神经网络的应用，大家可以在老师的指导下分成几个小组完成相关操作。首先在一张空白纸上写下一个随机的数字（0~9），然后拍照，把照片上传至计算机中，并将图片处理成28*28像素（利用Windows系统自带的图片处理工具进行处理），然后运行程序。在程序运行过程中，修改相关参数，如迭代次数、样本数量、学习率等，观察程序的运行结果。请参照表3.3.6完成参数设定并记录运行结果。各小组在完成体验后相互分享操作过程中遇到的问题及解决思路。

BP（Back Propagation）神经网络中的学习率通常取 0 和 1 之间的值，学习率越小，学习会越精细，但如果学习率太小，学习速度会很慢，若学习率太大，则学习效果会变差。

表 3.3.6　BP神经网络参数表与输出结果

序号	迭代次数	样本数量	学习率	识别测试数字	准确率
1	10	10	0.1	3	95%

拓展知识

BP神经网络是一种误差反向传播算法的学习过程，由信息的正向传播和误差的反向传播两个过程组成，其实现原理如图3.3.7所示。输入层各神经元负责接收来自外界的输入信息，并传递给中间层各神经元。中间层是内部信息处理层，负责信息变换，根据信息变换能力的需求，中间层可以设计为单隐藏层或者多隐藏层结构。最后一个隐藏层传递到输出层各神经元的信息，经过进一步处理后，完成一次学习的正向传播处理过程，由输出层向外界输出信息处理结果。当实际输出与期望输出不符时，进入误差的反向传播阶段。误差通过输出层后用一种特定的方法修正各层权值，向隐藏层、输入层逐层反传。然后重复这个过程，不断调整各层权值，直到网络输出的误差减少到可以接受的程度，或者达到预先设定的学习次数时结束这个过程。这就是BP神经网络学习训练的过程。

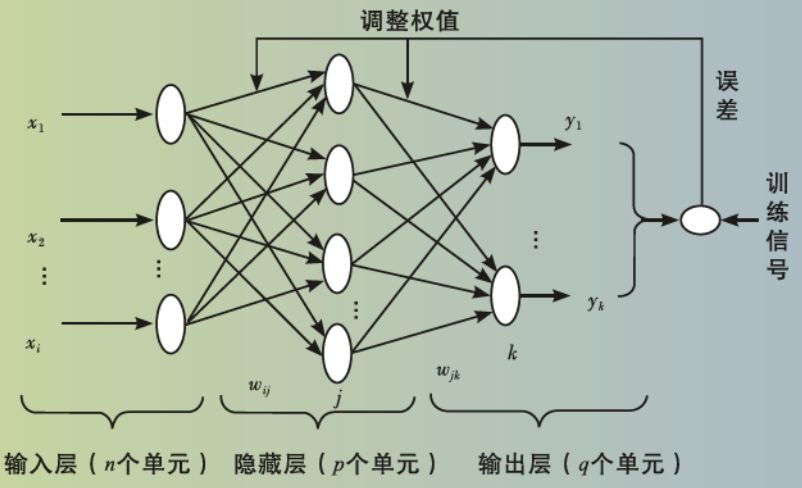


图 3.3.7　三层BP神经网络结构图

单元学习评价

通过本单元的学习，我们初步了解了机器学习的基本方法，知道了人工神经网络与深度学习等人工智能关键技术的应用场景。请根据所学知识，展开调查，深入理解机器学习的内涵，搜索一些采用机器学习实现的人工智能应用案例及其实现过程，回答以下问题。

1. 机器学习与人类学习有哪些相同之处与不同之处？

2. 针对你所搜索的某个通过机器学习实现的人工智能应用案例，对案例进行评价，重点评价技术方案的可行性与应用价值，并思考该应用是否能用机器学习方法以外的技术来实现，填写以下案例评估报告。

机器学习案例评估报告

【报告内容】

案例名称：_____

机器学习方法简介：_____

主要解决的问题：_____

应用价值：_____

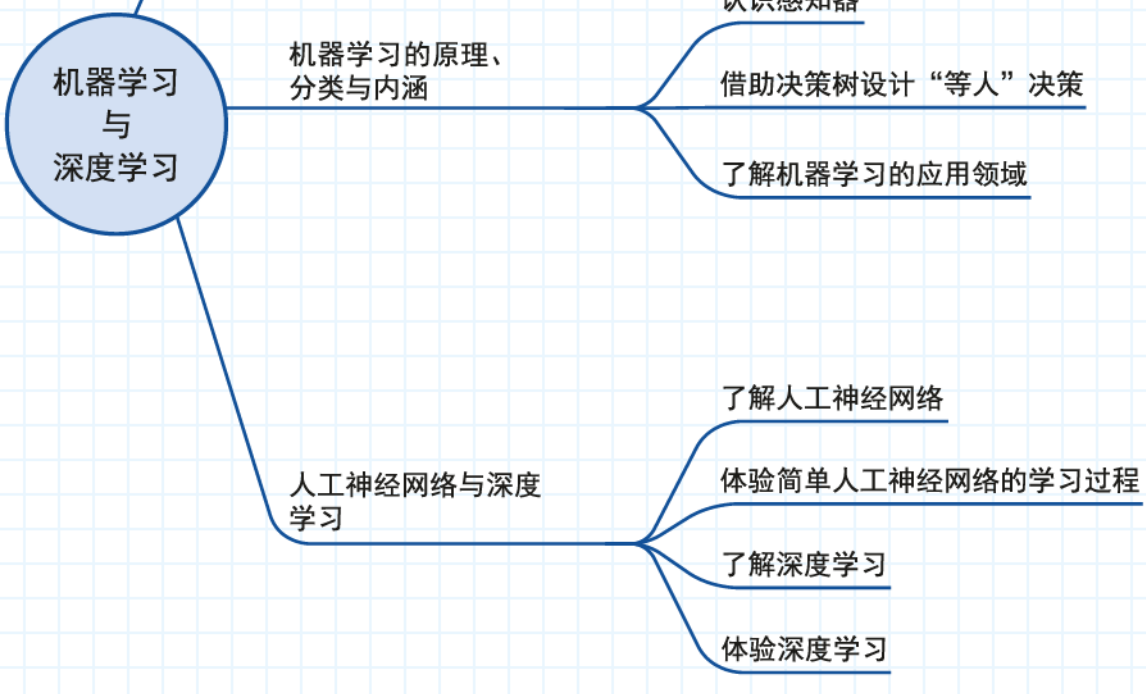
是否有替代技术方案：_____

【评价要点】

报告完整性：_____

报告内容合理性：_____

单元学习总结


机器学习
与
深度学习
机器学习的起源与发展
体验机器学习
了解机器学习的发展历程
机器学习的原理、
分类与内涵
认识感知器
借助决策树设计“等人”决策
了解机器学习的应用领域
人工神经网络与深度
学习
了解人工神经网络
体验简单人工神经网络的学习过程
了解深度学习
体验深度学习

第 4 单元　体验人工智能应用

斯坦福大学的专家们在研究报告《2030 年的人工智能与生活（AI and Life in 2030）》中断言：“现在到2030年人工智能可能会出现越来越有用的应用，有可能给我们的社会和经济带来深远的积极影响。”

人工智能在过去的60多年间取得了长足的发展。在较早的时候，人工智能还仅限于学术研究；21世纪，人工智能已经走进人们的日常生活。例如，计算机视觉和人工智能规划推动了视频游戏的发展，并一举超越了好莱坞电影产业；深度学习让语音识别走进每个人的手机，还被广泛应用到其他许多基于模式识别技术的日常应用中；自然语言处理（NLP）、知识表示及推理能让机器在智力竞赛节目中战胜人类冠军，并为Web搜索带来新力量。近年来，随着各种开源人工智能应用框架及人工智能开放平台的使用，人工智能技术变得更加友好，更多的专业技术人员和爱好者进入人工智能领域，并针对实际问题定制各类个性化的解决方案。

主流的人工智能应用框架，如TensorFlow、Caffe、OpenCog和Theano等，以及百度、腾讯、阿里巴巴和科大讯飞等的人工智能开放平台（也称AI开放平台）的出现，加快了人工智能技术在日常生活中应用普及的速度，拓宽了人工智能的应用场景。

在前面的单元中，我们通过用TensorFlow识别手写体数字，了解人工智能深度学习的基本原理，体验了腾讯与百度的人工智能开放平台中有关人脸检测、图片识别和语音识别等的人工智能典型应用。在本单元中，我们将围绕“搭建简单的人工智能应用模块”项目开展学习，通过数字化学习与动手实践，进一步了解开源人工智能应用框架和开放平台的功能与特点，学会基本的人工智能开发环境的搭建，初步实现人工智能技术在文本挖掘、字符识别、人脸识别和语音识别等方面的简单应用。

为了完成该项目，需要思考以下问题：人工智能开放平台有哪些？这些平台的基本工作流程与功能是怎样的？如何搭建基于人工智能开放平台的开发环境？怎样实现简单的人工智能应用？为此，我们将顺次完成以下任务：

- ◆ 通过实践了解人工智能开放平台，掌握其开发环境的搭建
- ◆ 通过词云创建及文本分类等操作，了解文本挖掘原理及其实现过程
- ◆ 实现中文手写体的识别并探讨其可能的创新应用
- ◆ 实现人脸识别并探讨其可能的创新应用
- ◆ 实现语音识别并探讨其可能的创新应用

4.1　体验人工智能开放平台

随着人工智能技术的迅速发展，人工智能应用渐渐融入人们的生活。本节我们将借助人工智能开放平台体验拍照识花，尝试搭建人工智能开放平台的开发环境，为后面具体的人工智能应用创新的实现打下基础。

学习目标

★ 体验拍照识花的基本过程与方法。

★ 学会搭建人工智能开放平台的开发环境。

★ 了解当前典型的人工智能开放平台。

近年来，在开源及开放思想的影响下，人工智能技术已经变得非常友好，技术门槛逐渐变低，越来越多的信息技术从业人员和爱好者开始涉猎人工智能领域，越来越多的人工智能应用产品来到了我们身边。

任务　通过实践了解人工智能开放平台并掌握其开发环境的搭建

※ 活动1　拍照识花

分小组登录百度 AI 开放平台，在老师的指导与帮助下，体验人工智能开放平台的应用。

第一步，从其“开放能力”菜单项中找到“图像技术”，再在其列表中找到“图像识别”，单击其下方的“植物识别”进入页面。

第二步，单击该页面下方“功能演示”区域的“本地上传”，将事先准备好的植物照片上传，如上传一张水仙花的照片，系统就会

把可能与图片一致的若干植物的名称按照可能性概率值的先后罗列在右上角部位。

参照上述步骤，各小组在相应的人工智能开发平台上选择不同的体验主题，开展人工智能应用上机实践体验与协作学习，然后以小组为单位分享体验结果，并将各组体验结果填入表 4.1.1 中。

表 4.1.1 “人工智能”在线体验记录表

平台名称	应用名称	上传的素材	在线体验结果
百度AI开放平台	拍照识花	水仙花照片	水仙的概率为0.96

人工智能开放平台上提供的产品服务，是通过提供API（Application Programming Interface，应用程序接口）或者SDK（Software Development Kit，软件开发工具包）的形式实现的，开发者只要通过调用服务，并根据实际需求不断调整参数设置，最终便可获得较满意的人工智能应用。由于SDK中已集成了大部分API，并编写了一套完整的API调用操作，所以借助SDK进行开发会非常便捷。利用人工智能开放平台搭建人工智能应用的一般流程如图4.1.1所示。

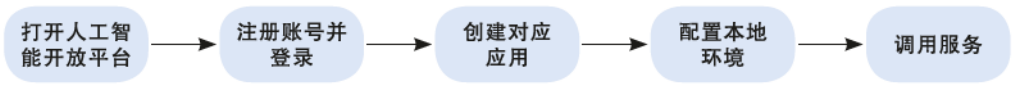


图 4.1.1 创建人工智能应用的一般流程

※ 活动2 搭建人工智能开放平台的开发环境

登录百度AI开放平台，在老师的指导与帮助下，搭建基于人工智能开放平台的开发环境。

第一步，在人工智能开放平台上注册账号并创建应用。

（1）注册账号并登录。

使用人工智能开放平台创建相关应用必须首先登录该平台。以百度AI开放平台为例，开发者成功注册百度账号后，便可通过账号登录到百度AI开放平台。

（2）创建应用并获取AppID等参数值。

在“管理控制台”页面中“产品服务”菜单的“人工智能”选项下面选择“语音识别”等需要的相应服务并“创建应用”，进入应用创建界面。填写应用名称与应用描述即可创建应用服务，此时平台会显示该应用的AppID、API Key以及Secret Key等参数，记下这些参数值备用。

第二步，配置本地开发与运行环境。

成功创建应用服务后，便可在平台上下载相应开发语言的SDK，下载并在本地机上安装后，便可在程序中直接调用SDK提供的相应接口和服务功能。Python 2.7以上版本自带了包管理工具pip，它提供了对Python 包的查找、下载、安装和卸载的功能。我们可以方便地使用pip自动下载并安装相应的SDK。

例如，下载并安装百度人工智能SDK，只需要在Windows的命令行窗口（可使用Windows键+R组合快捷键运行cmd命令打开）中输入“pip install baidu-aip”并按回车键，即可自动完成相应SDK包的搜索、下载和安装，如图4.1.2所示。

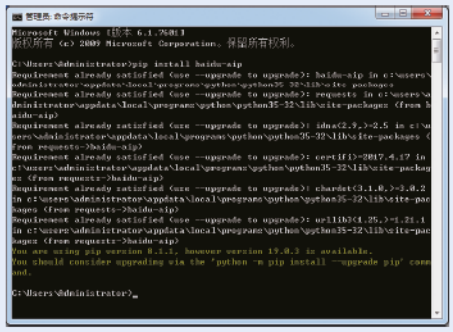


图 4.1.2　下载并安装baidu-aip SDK

接下来，可视情况尝试参照表4.1.2的内容搭建其他人工智能开放平台的开发环境，并将结果填入表4.1.2中。

表 4.1.2　搭建人工智能开放平台的开发环境活动记录表

平台名称	创建应用	相关授权码	SDK安装
百度AI开放平台	语音识别	AppID=15691890 APIKey=h8qM4kbwXYtHwdd0VEaSt1OM SecretKey=PnBPDDg7AkeXxmhImoMXGF4y3GsXOocy	安装成功
腾讯AI开放平台			
阿里AI开放平台			
讯飞AI开放平台			

国内四大人工智能开放平台

科技部在2017年召开了新一代人工智能发展规划暨重大科技项目启动会，会上公布了首批国家新一代人工智能开放创新平台名单，提出依托百度公司建设自动驾驶国家新一代人工智能开放创新平台，依托阿里云公司建设城市大脑国家新一代人工智能开放创新平台，依托腾讯公司

建设医疗影像国家新一代人工智能开放创新平台，依托科大讯飞公司建设智能语音国家新一代人工智能开放创新平台。

人工智能开放平台以API或SDK方式向开发者提供人工智能产品开发技术支持服务，开发者与平台合作不仅可以享受到人工智能开放平台带来的便利，而且能够推动相关领域人工智能技术的发展。从整体上看，上下游互动的结果使该行业的发展生态越来越完善。

● 百度AI开放平台

百度AI开放平台属于综合性开放平台，为开发者提供了全球前沿水平的语音、图像、自然语言处理、知识图谱等多项人工智能技术的产品服务，提供了开放对话式人工智能系统、智能驾驶系统两大行业生态以及智能零售、智能政务等行业的解决方案。

● 腾讯AI开放平台

腾讯AI开放平台依托腾讯AI Lab、腾讯优图、WeChat AI等实验室，汇聚腾讯的AI技术能力，开放100余项AI功能接口供行业使用。线下还通过AI加速器帮助和扶持AI创业者，打造AI开放新生态。腾讯发布了一款AI医学影像产品——腾讯觅影。腾讯觅影是首款AI食管癌筛查系统，检测准确率已超过90%；在肺结节方面，腾讯觅影可以检测出3毫米及以上的微小结节，检测准确率已超过95%。未来腾讯觅影将与医学院校或医疗机构合作，助力更多病种检测。

● 阿里AI开放平台

阿里云ET城市大脑是目前全球最大规模的人工智能公共系统，可以对整个城市进行全局实时分析。目前阿里云ET城市大脑已经在杭州、苏州等地落地。据2017年新闻报道，城市大脑接管了杭州128个信号灯路口，试点区域通行时间减少15.3%，高架道路出行时间节省4.6分钟；在杭州主城区，城市大脑日均事件报警500次以上，准确率达92%；在杭州另一区，120救护车到达现场时间缩短一半。

● 讯飞AI开放平台

讯飞AI开放平台属于专业性开放平台，是科大讯飞推出的以语音交互技术为核心的人工智能开放平台，为开发者提供语音识别、语音合成等语音技术SDK，其语音识别的准确率可达到98%以上。此外，讯飞AI开放平台还提供智能家居、智慧金融、智慧教育等行业解决方案。

4.2　体验中文文本挖掘

文本挖掘指的是从文本数据中获取有价值的信息和知识的技术，是数据挖掘技术的重要组成部分，是人工智能的一个重要应用领域。实现文本挖掘的技术细节易于理解，也很有代表性。本节我们将通过词云图绘制以及文本分类等文本挖掘应用的实现，来了解文本挖掘的原理，体验文本挖掘的魅力，初窥人工智能技术的部分细节。

学习目标

★ 通过文本挖掘应用实例，了解文本挖掘的实现原理与常用方法。

★ 了解文本挖掘的常用算法与工具，通过相关活动体验简单文本挖掘的实现过程。

腾讯的写稿机器人“梦幻写手”早就能在国家统计局公布年度居民消费价格指数（CPI）的第一时间“写”出新闻稿。近两年，微软的“小冰”也越来越厉害了，诗词歌赋似乎“样样精通”。我们从小努力学习，知道看懂一篇文章并写出有质量的文章，是需要从识字开始历经数年才能习得的技能，那么人工智能究竟是如何做到看懂文章并完成撰写新闻稿工作的呢？

任务　通过词云创建及文本分类等操作，了解文本挖掘原理及其实现过程

※ 活动1　试用好玩的“图悦”软件

“图悦”是一个支持生成词云图的在线工具，能自动统计词频、制作词云图并导出词频信息表，如图4.2.1所示。

图4.2.1 “图悦”的主界面

词云图是对文档中的关键词按其重要性采用大小不同的方式予以视觉上突出显示的图形，以方便浏览者快速领略文档的主旨。使用“图悦”制作词云图的步骤如下。

第一步，打开“图悦”界面，将需要分析的文本直接粘贴在左边栏的“待分析长文本或URL”中。

第二步，单击右上角的“分析出图”按钮。

第三步，观察右边栏中词云图的变化，图中文字大小反映了输入文本的词频信息，即该词在文本中出现次数的信息，个头最大的词是词频最高的词，随着词频由高到低的变化，词的个头逐渐变小。

第四步，导出词频信息表。用文字个头大小描述词频虽然形象，但不够精确，可以单击“图悦”界面最右边的“导出Excel”按钮，将词频的统计结果导出到Excel文件中。对应的词频信息表如图4.2.2所示。

	A	B
1	关键词	词频
2	人事	25
3	员工	19
4	人力资源部	10
5	反馈	5
6	命题	4
7	制度	4
8	技巧	4
9	环境	4
10	晋升	3
11	人力	3

图4.2.2 词频信息表

“图悦”支持最大长度为100万汉字的纯文本，最多可显示排位前150位的关键词。选择一本你喜欢的小说（字数控制在100万以内），借助“图悦”给出的词云图对小说人物或事件进行分析；也可以对小说不同章节进行处理，根据结果分析其侧重点。教科书配套资源中为大家提供了部分供测试的文档，大家也可以上网寻找自己感兴趣的文档进行分析。

词频率

词频率（Term Frequency，TF），衡量一个词在文档中出现的频繁度，通常表示为：TF(t)=（词条出现的次数）/（文档中词的总数）。

※ 活动2　基于Python库实现中文分词，了解中文分词原理

体验了“图悦”制作词云图的过程，一个问题便浮现出来：“图悦”是如何做成这件事的？看似好玩的词云图生成，其实涉及了一些文本挖掘的基础知识。接下来，我们将在Python环境中一步步实现词云图的制作并了解其实现的原理。

文本挖掘

文本挖掘是指从大量文本数据中抽取事先未知的、可理解的、最终可用的信息或知识的过程。文本挖掘被广泛应用于搜索和信息检索、文本聚类、文本分类、Web挖掘、信息抽取和自然语言处理等方面。

词袋模型

词袋模型（bag-of-words model）是用于描述文本特征的一个简单的数学模型，是一种常用的文本特征提取方式。词袋模型将文档看作一个“装有若干词语的袋子”，在词袋模型中只考虑词语在文档中出现的次数，而忽略词语的顺序以及句子的结构等因素。

例如，以下面的示例文档为待分析对象：

小明喜欢打羽毛球，也喜欢打篮球。

我们可以将上述文本表示为一个由词语及其出现次数组成的二元组的集合：

{（小明：1）（喜欢：2）（打：2）（羽毛球：1）（也：1）（篮球：1）}

这个集合就是这段文字对应的“词袋”。

词袋模型对文档进行了很大程度的简化，但一定程度上仍然保留了文档的主题信息。根据词袋中的“小明”“羽毛球”“篮球”等词语，仍然可以知道这个文档与小明以及两个可能的体育主题相关。词袋模型的基本思想就是忽略不重要的词句结构，保留并统计体现主题的词语。

有了词袋之后，可以构造一个包含若干词语的“词典”（vocabulary），并借助这个词典将词袋转换为用于表征该文档特征的特征向量，如词计数向量（term counting vector）或词频向量（term frequency vector）等。词频的原义指一个词在文档中出现的次数，为计算归一化，我们把词频除以句子中词的总数的结果也称为词频。

例如，我们可以构造一个包含六个词语的词典，如表4.2.1所示。

在数学中，向量指具有大小和方向的量，它可以形象化地表示为带箭头的线段。

例如，文中的词计数向量值就是向量的大小，每个值对应的属性就代表向量的方向。

如果词典的大小为 V，对其中每一个词语，我们统计其在所有文档中出现的总数 n_i，再除以文档的总词语数 n，就可以得到对应的词频 x_i，我们再将所有的词频组合在一起，就可以得到一个维数为 V 的词频向量 $f=(x_1, x_2, \cdots, x_V)$。

表4.2.1 词典所包含的六个词语

序号	1	2	3	4	5	6
词语	小明	喜欢	打	羽毛球	也	篮球

将每个词语在文档中出现的次数按照词语序号排列起来，就得到这个文档的词计数向量$n=(1, 2, 2, 1, 1, 1)$。

我们还可以对词计数向量进行归一化（即缩放向量的长度，使所有向量长度之和为1），得到词频向量$f=(\frac{1}{8}, \frac{1}{4}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8})$。

通常并不要求词典包含文本中出现过的所有词语。例如，如果使用下面这个只包含四个词语的词典（如表4.2.2所示），示例文档的词计数向量和词频向量就分别变为$n=(1, 2, 1, 1)$与$f=(\frac{1}{5}, \frac{2}{5}, \frac{1}{5}, \frac{1}{5})$。

表4.2.2 词典所包含的四个词语

序号	1	2	3	4
词语	小明	喜欢	羽毛球	篮球

在实际应用中，待分析对象往往包含多个甚至是许多文档，可以使用一个公共的词典对其中的所有文档进行词频统计。下面以包含以下三个文档的待分析文档集为例进行分析。

文档1：小明喜欢打羽毛球，也喜欢打篮球。

文档2：小明去公园放风筝。

文档3：小明的学校开设了人工智能课程。

首先，我们从文档中提取所有出现过的词语，形成一个词典，如表4.2.3所示。

表4.2.3 语料词典

序号	1	2	3	4
词语	小明	喜欢	打	羽毛球
序号	5	6	7	8
词语	也	篮球	去	公园
序号	9	10	11	12
词语	放	风筝	的	学校

续表

序号	13	14	15	16
词语	开设	了	人工智能	课程

接下来，我们统计每个文档中每个词语出现的次数，如表 4.2.4 所示。

表 4.2.4　文档词语统计表

词语	小明	喜欢	打	羽毛球	也	篮球	去	公园	放	风筝	的	学校	开设	了	人工智能	课程
小明喜欢打羽毛球，也喜欢打篮球。	1	2	2	1	1	1	0	0	0	0	0	0	0	0	0	0
小明去公园放风筝。	1	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0
小明的学校开设了人工智能课程。	1	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1

统计结果即三个文档的词计数向量，分别如下所示：

n_1=（1，2，2，1，1，1，0，0，0，0，0，0，0，0，0，0）

n_2=（1，0，0，0，0，0，1，1，1，1，0，0，0，0，0，0）

n_3=（1，0，0，0，0，0，0，0，0，0，1，1，1，1，1，1）

词袋模型相对比较简单，在实际使用过程中，还需要与一些文本处理技术相搭配，才能取得较好的效果。图4.2.3展示了利用词袋模型构建文本特征的基本流程。

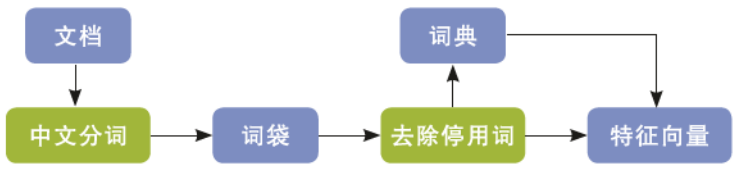


图 4.2.3　词袋模型应用流程图

中文分词

从词袋模型应用流程图可以看出，在构建词袋之前，需要先将句子中的词语分开。对于英文来说，这个过程是非常容易的，只需要以空格和标点符号为依据，就可以将其中所有的词语分隔开来。但在中文文本中，所有的字词连接在一起，计算机并不知道一个字应该与其前或后的哪一个字或几个字连成词语，还是应该自己形成一个词语。因此，在构建词袋之前，需要先借助中文分词技术将文本中的词语分开。中文分词一般是基于统计的方法来实现。

分词是中文自然语言处理技术中最基本、最底层的部分，分词精度对后续技术模块影响很大。一般情况下，为了提高精度，专业化的大型

中文文本分类系统常常定制开发自己的分词系统。例如，北京理工大学开发的中文分词系统、哈尔滨工业大学的语言云系统等。这类分词系统的精度、完整性和稳定性较好。 jieba分词是SunJunyi开发的一款Python中文分词组件，作为Python的外部库，其占用资源较小，对常识类文档的分词精度较高。

在Windows命令窗口运行“pip install jieba”就可以完成jieba的安装。jieba支持以下三种分词模式。

（1）精确模式 jieba.cut(str)：将句子精确切分，适合文本分析。

（2）全模式 jieba.cut(str, cut_all=True)：把句子中所有可以成词的词语都扫描出来，速度非常快，但是不能解决有歧义的问题。

（3）搜索引擎模式 jieba.cut_for_search(str, cut_all=True)：在精确模式的基础上，对长词再次切分，适用于搜索引擎分词。

例如，在Python命令行窗口中运行下述命令即可完成对应用文本的分词：

```
import jieba          #导入jieba分词库
jieba.lcut("小明喜欢打羽毛球，也喜欢打篮球。")  #用“lcut”可以直接输出分词列表
```

运行结果：

['小明', '喜欢', '打', '羽毛球', '，', '也', '喜欢', '打篮球', '。']

停用词

上面几个例子的词典中包含了一些诸如“的”“也”“了”的词，这些词出现次数很多，但对判断文档主题没什么意义，类似这种不携带任何主题信息的高频词常常被当作停用词（stop word）删除，常见的停用词还有“只”“当”和“从”等连词或介词。

在构建词典时，我们通常还需要将出现次数极低的低频词删除。这类词语通常是一些不常用的专有名词，如果我们过度依赖这些词语对文章的主题进行归类，就有可能出现因过度拟合导致失真的现象。而且，如果我们将这些低频词全部收录到词典中，将大大增加词典的大小以及特征向量的维数，给计算带来困难。

去除停用词可以用下列命令实现。

新列表=[word for word in 源列表 if word not in 停用词列表]，这个命令的使用前提是停用词要在分词过程中被正确拆分。

例如，要去除前面例子中的停用词“也”以及标点符号，可以通过下列命令实现：

```
print([w for w in jieba.cut("小明喜欢打羽毛球，也喜欢打篮球。")if
```

```
w not in ('也',', ',',。')])
```

运行结果：

['小明','喜欢','打','羽毛球','喜欢','打篮球']

我们会发现，“也”“。”以及“，”都作为停用词被去除了。对于长文档，比如《红楼梦》要去除的停用词会有成百上千个之多，通常的做法是建立一个有针对性的“停用词”文档，在使用时，直接导入该文档，再用上述命令将停用词去除。

大家可以试着对其他文本内容进行分词，并去除停用词。如果同学们感兴趣，可以自己尝试上网搜索如何使用停用词库的知识。

词频率与逆文档频率

通过上述方法去除停用词后，实际上还有许多高频词依然对文档主题没有什么说明作用，即有些词的词频虽然比较高，但并不能在说明文档主题倾向方面产生作用。经验证明，仅通过“词频”这个指标衡量词语的重要性是不够的。在文本挖掘中通常采用“词频率与逆文档频率（TF-IDF）”指标衡量文档中词语的重要性。其中，逆文档频率（Inverse Document Frequency，IDF）用于模拟在该待分析文本的实际使用环境中某一个词的重要性（出现最多的词不一定就是最重要的），算式为IDF(t) = log（待分析文档总数 / 含有该词的文档总数），为了避免出现分母为0的情况，有时也会将分母加1。

词频率与逆文档频率表示为：TF-IDF = TF*IDF。TF-IDF值越大，则这个词成为关键词的概率就越大，即权重越高。从词频率与逆文档频率计算公式中可以看出，TF-IDF与一个词在所有文档中的出现次数成正比，与含有该词的文档数成反比。换句话说，一个词语在所有文档中出现次数越多，同时包含该词的文档数越少，越能够显示该词的重要性。

以前面的三篇文档为例，“小明”和“羽毛球”的词向量表如表4.2.5所示。

表 4.2.5　词向量表

文档中的词	包含该词的文档数	TF	IDF	TF-IDF
小明	3	0.15	0	0
羽毛球	1	0.05	0.477	0.024
喜欢				

从词向量表可以看出，1号词“小明”在三篇文档中均出现了，

则“小明”一词的词频率为3/20，逆文档频率为$\log_2$（3/3）=0，其TF-IDF值也为0。再看“羽毛球”一词，它只在一篇文档中出现，它的词频率为1/20，逆文档频率为$\log_2$（3/1）≈0.477，TF-IDF值约为0.024。这与我们的直观想法相符，即“小明”一词在说明文本主题意义方面的重要性要低于“羽毛球”一词。

请把表4.2.5补充完整。

TF-IDF是对简单依赖词频的一种修正，可以更好地捕捉文本中的重要信息。在文本挖掘中，通常都用TF-IDF值来说明文档特征。jieba就是通过TF-IDF算法实现关键词统计的。

jieba中的TF-IDF算法的调用命令如下：

jieba.analyse.extract_tags(sentence，topK，withWeight，allowPOS)其相关参数的说明如下。

sentence：待提取的文本。

topK=20：返回几个TF-IDF权重最大的关键词，这里指返回前20个权重最大的关键词。

withWeight=False：是否一并返回关键词权重值，取值为False指不返回，取值为True则返回。

allowPOS=()：仅包括指定词性的词，默认值为空，即不筛选 。

在使用该命令时，除了导入jieba分词包，还需要导入analyse包。

例如，jieba.analyse.extract_tags("小明喜欢打羽毛球，也喜欢打篮球。",withWeight=True)的运行结果是：

[('打篮球', 2.34069061492),
('喜欢', 2.281035361208),
('小明', 2.22561778594),
('羽毛球', 2.07486342548)]

上述运行结果呈现的就是关键词与对应的TF-IDF值，可以看到该运行结果与前面的计算结果并不一致，原因是jieba在计算时应用了它内部默认的词典与计算模型，并非表4.2.3中的词典。

学习了分词、去停用词及TF-IDF算法，接下来我们就可以绘制词云图了。Python提供了比较标准的词云生成功能，主要用到其中的wordcloud和matplotlib两个库。wordcloud库主要用于生成词云，matplotlib库用于绘图，可以将词云以图的形式展示出来。wordcloud和matplotlib同样也可以用pip install命令来安装。安装完成并通过“import”命令导入库后，就可以通过命令“wordcloud().generate(text)”和“matplotlib.pyplot.imshow()”实现词云图的绘制了。

例如，通过以下两条命令就可以画出如图4.2.4所示的词云图。需要注意的是，“图悦”给出的词云图仅是依据TF而来的，我们绘制的词云图是基于TF-IDF的，能更准确地说明关键词的重要性。

“cloudobj=wordcloud.WordCloud(font_path=myfont).generate('小明喜欢 打 羽毛球 ， 也 喜欢 打 篮球。')”#括号中的文本是事先完成分词的

“matplotlib.pyplot.imshow(cloudobj)”#绘制词云图

图 4.2.4　词云图

大家可能觉得奇怪，似乎我们并没有给出关键词权重的计算方法和去除停用词的过程。实际上，这些过程都已经集成在wordcloud库中了，但其自带的停用词表不是很完善，对于要求比较高的文本挖掘工作来说，一般都会自定义停用词表或去相关网站下载指定的停用词表。

请尝试用上述方法绘制《红楼梦》（可以自行下载相应的文本文档）第一回的词云图，也可以利用教科书配套资源中提供的其他文档绘制词云图，或者从网上搜索感兴趣的文档进行练习。

文本相似度

在数学课中我们学过余弦定理，但是同学们一定想不到余弦定理能帮助人们完成复杂的文本挖掘工作，如新闻分类。谷歌的“新闻”服务就是从采用余弦定理开始完成新闻分类的，从而节约了大量的人力资本。

我们用每个文档中词语的TF-IDF来组成该文档的特征向量，并采用余弦定理对其进行计算。如向量$a=(x_1, x_2)$，向量$b=(y_1, y_2)$，则$\cos\theta=\dfrac{x_1\times y_1+x_2\times y_2}{\sqrt{x_1^2+x_2^2}\sqrt{y_1^2+y_2^2}}$，当向量维数大于2时，如$a=(x_1, x_2,\cdots, x_n)$，$b=(y_1, y_2,\cdots, y_n)$，$\cos\theta=\dfrac{\sum_{i=1}^{n}x_i\times y_i}{\sqrt{\sum_{i=1}^{n}x_i^2}\sqrt{\sum_{i=1}^{n}y_i^2}}$，计算得到的余弦值越接近1，则两个文本越相似。

例如，一个文本的特征向量$v_1=(1, 2, 3)$，而另一个文本的特征向量是$v_2=(4, 7, 5)$，则它们的相似度可表示为

$$\cos\theta=\frac{1\times4+2\times7+3\times5}{\sqrt{1^2+2^2+3^2}\sqrt{4^2+7^2+5^2}}=0.92966968$$

※ 活动3 了解文本相似度的计算方法

请将上述文本相似度的计算过程转换成Python程序，并将下列代码补充完整。

```
import  math
from  numpy import *
v1 = mat([1,2,3])                    #将列表变为numpy向量的形式
v2 = mat([4,7,5])                    #将列表变为numpy向量的形式

Lv1 = math.sqrt(v1*v1.T)             #计算 √(∑ x_i^2)，v1.T是v1的转置矩阵

Lv2 =_____________________           #计算 √(∑ y_i^2)
cosV12 = v1*v2.T/(Lv1*Lv2)
print(cosV12)
```

同学们看懂这段程序了吗？可以尝试在Python环境中边运行代码边理解。由于涉及向量计算，这里用到了numpy模块。大家可以将v1和v2替换成其他文本的特征向量，并计算它们的相似性。

上述方法是从微观层面通过特征向量之间的余弦值计算实现文本的相似性比较。实际上，我们还可以通过机器学习框架完成更复杂的文本分类工作。

朴素贝叶斯算法

朴素贝叶斯算法是文本分类中比较常用的一种方法，它的算法原理可以由联合概率公式变换得来。联合概率公式可表示为：

$P(Y,X)=P(Y|X)*P(X)$
$=P(X|Y)*P(Y)$

其中，P(Y,X) 指X和Y同时出现的联合概率，P(X) 和 P(Y) 分别指X出现的概率和Y出现的概率，P(Y|X) 指满足X出现的条件下Y出现的概率，P(X|Y) 指满足Y出现的条件下X出现的概率。因此联合概率公式等于X出现的概率乘当X出现的条件下Y出现的概率，也等于Y出现的概率乘当Y出现的条件下X出现的概率。将上述公式稍作变化即可得出贝叶斯公式：

$$P(Y|X)=\frac{P(X|Y)*P(Y)}{P(X)}$$

文本分类

文本分类是文本挖掘的一个子类别，它对片段、段落或文件进行分组和归类，在使用数据挖掘分类方法的基础上，标记出某个文本是哪一种类型的。文本分类是文本挖掘最基础也是最广泛的核心技术，被广泛应用于文本检索、垃圾邮件过滤及文本信息优先级评估等领域。文本分类包含以下几个基本步骤。

第一步，文本有效信息的提取。包括文本预处理，如分词、去除停用词及相关文档清理等工作；其次是特征抽取，即从文档中抽取出反映文档主题的特征向量。

第二步，选择朴素贝叶斯算法进行模型训练。这里我们选择基于词袋模型的文本分类算法。

第三步，选择Sklearn作为文本分类工具包。

第四步，对分类结果进行评价与反馈。根据分类结果的准确率对模型进行评估，调整参数设置，不断优化模型，优化分类结果。

Sklearn简介

Sklearn（Scikit-learn）是一个开源的Python语言机器学习工具包，它涵盖并实现了几乎所有主流机器学习算法，针对从数据预处理到训练模型的各个方面，包括数据降维、回归算法、分类算法、聚类分析等，并且提供了一致的调用接口。Sklearn基于Numpy等Python计算库，提供了高效的算法实现，具有文档齐全、接口易用以及算法全面等优点。在实际应用中，使用Sklearn可以极大地节省编写代码的时间，使人们有更多的精力去分析数据、调整模型和修改参数以达到更好的效果。与前面学习的TensorFlow深度学习框架相比，Sklearn属于通用机器学习库，有时也可以将Sklearn和TensorFlow等多种应用框架结合起来使用。

※ 活动4　体验基于Sklearn的文本分类实现过程

请根据所学知识将《红楼梦》第一回和第二回中的所有段落按章节进行分类，并测试分类准确率。根据文本分类的基本步骤，写出实现思路及主要实现代码，并画出流程图。

操作过程提示：

（1）将《红楼梦》前两回的内容提出另存成一个文档备用，建议将字数不足50字的段落去除。

（2）用jieba进行分词预处理。

（3）导入sklearn.feature_extraction库进行文本特征提取，并将数据划分为训练集和测试集。

代码提示：

```
from sklearn.feature_extraction.text import CountVectorizer
countvec=CountVectorizer()          #导入特征提取库，提取文本特征
#导入训练集和测试集分类器
wordmtx=countvec.fit_transform（预处理后的文本）
from sklearn.model_selection import train_test_split
#将提取特征后的数据划分为训练集和测试集，并将训练集和测试集按3：7的比例划分
x_train,x_test,y_train,y_test=train_test_split(wordmtx,
test_size=0.3,random_state=111)
```

（4）导入贝叶斯模型库拟合贝叶斯模型，对文档进行分类。

代码提示：

```
from sklearn import naive_bayes      #导入贝叶斯模型库
NBmodel=naive_bayes.MultinomialNB()
NBmodel.fit(x_train,y_train)     #拟合朴素贝叶斯模型
NBmodel.predict(x_test)          #验证集预测
```

（5）对模型进行评估，预测文档分类的准确率。

代码提示：

```
#对训练集和验证集结果的正确率打分
print('训练集：',NBmodel.score(x_train,y_train),
    ',验证集：',NBmodel.score(x_test,y_test))
```

在本活动的实现过程中，我们主要采用了前面学过的基于词袋模型的特征抽取法。Sklearn有两个类 CountVectorizer和TfidfVectorizer，分别对应词频统计和TF-IDF两种不同的文本特征提取，大家可以对比两种分类结果有什么不同。源程序见教科书配套资源。

同学们可以利用教科书配套资源中提供的文档进行分类练习，也可以上网搜索自己感兴趣的文档并做适当处理后进行分类练习。

4.3　字符识别及其应用

在人工智能领域，字符识别是机器学习模式识别应用的一种，其实现是一个非常复杂的过程，如第3单元借助TensorFlow深度学习框架实现手写体数字的匹配识别就是一个典型的技术实现案例。为了促进应用，降低技术实现门槛，许多人工智能开放平台将字符识别技术实现的过程进行了封装，以API接口的形式对外提供服务。开发者无须关心具体识别的技术实现过程，只需要将待识别的图像提交便可从反馈信息中得到识别结果，这会极大地促进字符识别技术在各行各业中的应用。

字符识别的全称是光学字符识别（Optical Character Recognition，OCR），是指对文本资料的图像文件进行分析与识别处理，获取文字及版面信息的过程。当前，字符识别在社会生活的不同行业中有着广泛的应用，包括通用文字识别、表格文字识别、身份证识别、银行卡识别、名片识别、驾驶证和行驶证识别、各类票据识别等。

本节我们将借助相关人工智能开放平台的API接口，尝试实现以中文手写体为代表的字符识别，并探索它在日常学习和生活中的应用。

学习目标

★ 尝试实现中文手写体识别。

★ 掌握借助人工智能开放平台的 API 接口实现具体应用的一般方法。

★ 探索基于字符识别等人工智能技术的创新应用。

任务 实现中文手写体的识别并探讨其可能的创新应用

※ 活动1 中文手写体识别的实现

百度、腾讯等的人工智能开放平台提供了非常丰富的字符识别服务API接口，通过调用这些接口，我们可以方便地实现对汉字等字符的识别。下面我们就以百度AI开放平台为例实现中文手写体的识别，体验借助人工智能开放平台的API接口实现人工智能应用的一般过程和方法。

在白纸上写上一些你想要识别的汉字（要求10个字符以上），拍照上传至计算机中（注意不要使照片上的文字颠倒或倾斜），分别以img1.jpg、img2.jpg、img3.jpg命名。

为简化程序的设计和运行，可以参照相应API帮助文档把相对复杂的平台API接口调用过程封装成类，以方便后续程序的重复调用。我们在Python IDLE中编写代码将手写体识别功能封装成类，并保存在module.py文件中。部分参考代码如下。

```
class Module():                                    #定义手写识别类
  def __init__(self, api_key, secret_key): #类初始化
    ......
  def get_file_content(self, file_path):       #以二进制方式读取图片
    ......
  def hand_write(self, file_path):             #实现手写体识别
    image = self.get_file_content(file_path)  #读取图片
    #将二进制数据转化为base64编码
    base64_image = base64.b64encode(image)
    #创建请求参数并赋值
    params = {'image': base64_image.decode('utf-8')}
    #生成请求url，参数为access_token
    url = "https://aip.baidubce.com/rest/2.0/ocr/v1/handwriting?
access_token=" + self.access_token
    #发送API请求，返回识别结果数据
    res = requests.post(url=url, headers={'Content-Type': 'application/
json'}, data=params)
    json = res.json()                              #将返回结果转化为json格式
    #返回json数据里的识别文字结果
    return json['words_result'][0]['words']
```

将img1.jpg、img2.jpg和img3.jpg与module.py和run.py置于同一目录下并运行run.py，分别识别img1.jpg、img2.jpg和img3.jpg这三张手写体图片，将识别过程记录在表4.3.1中，并对结果进行讨论。

表 4.3.1　手写体识别过程记录表

图片文件名称	图片上的手写文字内容	识别结果	正确率
img1.jpg			
img2.jpg			
img3.jpg			

根据识别结果，试分析影响其识别正确性的因素有哪些，并结合所学人工智能知识和技术，讨论可采取哪些措施不断提升识别正确率。

可根据识别场景的不同将字符识别分为识别特定场景的专用OCR和识别多种场景的通用OCR。身份证识别、车牌识别、银行卡识别、火车票识别等都属于专用OCR的典型实例。通用OCR可以用于更复杂的场景，也具有更大的应用潜力。但通用OCR的应用场景不固定，文字布局多样，因此难度更高。

典型的OCR的技术路线如图4.3.1所示。

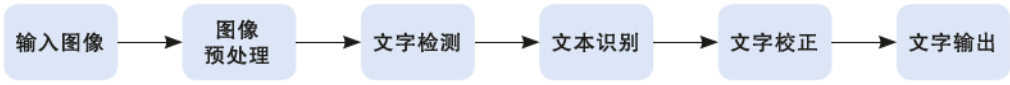


图 4.3.1　典型的OCR的技术路线图

在图4.3.1中，图像预处理通常是针对图像的成像问题进行修正，常见的预处理过程包括几何变换（透视、扭曲、旋转等）、畸变校正、去除模糊、图像增强和光线校正等。文字检测即检测文本的所在位置和范围及其布局，通常也包括版面分析和文字行检测等，它主要解决哪里有文字、文字在图片上的范围有多大的问题。而文本识别则是在文字检测的基础上，对文本内容进行识别，将图像中的文本像素信息转化为文本字符信息。识别出的文本字符通常需要再次核对、校正以保证其正确性，因此文字校正也是OCR一个非常重要的环节。

在文本识别过程中，当识别的内容是由词库中的词汇组成时，称为有词典识别（Lexicon-based），反之则为无词典识别（Lexicon-free）。在字符识别过程中，对识别准确率影响最大的技术瓶颈一般是文字检测和文本识别这两个环节，它们对人工智能技术水平的要求也最高。

※ 活动2 智能阅卷与字符识别应用创新探索

手写体识别技术在社会生活的各个领域有着广阔的应用前景，从传统文稿的智能整理分析到智能阅卷或智能批改，都有它大显身手的空间。下面我们以某论述题的智能批改为例，开展手写体识别应用的创新探索。

◆ 应用分析

一般论述题都有相应的答题要点以及各自的赋分分值，每个答题要点又包含一个或若干个关键词。批改过程中，一般根据答题结果是否切中对应答题要点决定是否给分。

◆ 实现思路

借助快速扫描仪获取答题图片；先对答题图片开展手写体识别获得答题文本；然后通过相应文本挖掘工具分析获得相应的关键字词；再对相应的关键字词与答题要点中的关键词进行匹配；最后根据匹配结果进行赋分。其实现参考流程图如图4.3.2所示。

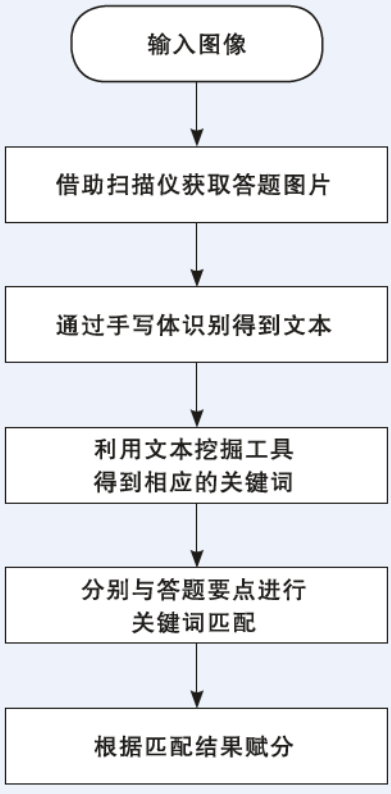


图4.3.2 智能批改技术实现流程图

结合前面的应用案例，你能实现这一应用过程吗?

借助OCR技术，可以快速地将图片上的文字内容转换成相应的文本，再对这些文本进行分析处理以满足应用需求。如对马路上的车牌进行OCR识别可快速锁定车辆信息；将手工填写的医疗记录表单自动识别为文字，即可方便医疗信息的电子化存储和分析；对快递单上的

寄件人和收件人信息精确识别可大大减轻人工输入的负担。可以说，OCR技术的应用提高了人们的工作效率，极大地方便了人们的工作和生活。

在日常生活中，哪些场景可以应用相关的OCR技术（如各类证件识别等），让人们的工作、生活更加方便？请将你的想法与创意记录在表4.3.2中。

表 4.3.2　OCR技术应用场景创意记录表

你创设或建议的场景	应用了哪些OCR技术	预计带来的便利

参照前面描述的智能批改的分析与实现过程，选择表4.3.2中你最熟悉或喜欢的一个应用场景，尝试将应用分析与创意实现过程表述出来。

4.4 人脸识别及其应用

人脸识别也称为人像识别或面部识别，是基于人的脸部特征信息进行身份识别的一种生物识别技术，其中包含借助摄像机等硬件直接采集图像或从视频流中抓取图像，再在图像中检测和跟踪人脸，进而根据检测到的人脸对其身份进行识别等环节。人脸识别的技术层面可依次分为人脸比对、人脸属性检测和人脸搜索，主要应用场景有安防监控、身份验证、人脸美化、智能相册、人脸3D建模和互动营销等。本节我们将借助人工智能开放平台API，通过在部分应用场景中的体验性尝试学习这些技术。

学习目标

- ★ 了解借助人工智能开放平台 API 实现人脸比对的原理与方法，体验技术实现过程。
- ★ 了解人脸识别实现的一般过程，开展人脸库管理与人脸搜索，实现人脸识别。
- ★ 探讨人脸识别在日常学习与生活中可能的应用思路。

上一节我们已经初步掌握了如何借助人工智能开放平台API接口实现手写体识别，那么，人脸比对、人脸检测和人脸搜索与之相比较又有哪些异同呢？让我们一起探究吧！

任务 实现人脸识别并探讨其可能的创新应用

※ 活动1 比对两张人脸的相似度

人脸比对是通过提取人脸的特征信息，计算两张人脸的相似度，给出相似度评分，从而实现对其是否为同一人的判断。在已知用户ID的情况下帮助确认是否为用户本人的比对操作，即一对一身份验证，

已应用于企业刷脸考勤、酒店自助入住、金融远程开户等诸多应用场景中。

准备若干用于比对的人脸照片1.jpg、2.jpg……，既有含不同人的照片，也有含同一人的不同照片。

参照百度AI开放平台有关的API技术文档，在Python的IDLE窗口中编写代码实现人脸照片的比对，部分参考代码如下。

```
class Face():       #创建人脸识别类
  def __init__(self, app_id, api_key, secret_key):       #初始化类
    #创建百度云客户端
    self.client = AipFace(app_id, api_key, secret_key)
  def match(self, file1_path, file2_path):          #定义人脸比对功能
    image1 = base64.b64encode(open(file1_path, 'rb').read()).
decode('utf-8')
    image2 = base64.b64encode(open(file2_path, 'rb').read()).
decode('utf-8')
    params = [{'image': image1, 'image_type': 'BASE64'},
              {'image': image2, 'image_type': 'BASE64'}]
    res = self.client.match(params)       #调用API接口返回比对结果
    return res['result']['score']         #提取出结果中的相似度分数
#初始化类实例
face = Face(app_id, api_key, secret_key)
#比对两张人脸图片，返回相似度
score = face.match('1.jpg', '2.jpg')
print('相似度为：', score)       #输出相似度分数
```

将程序文件和需要比对的1.jpg和2.jpg两张图片文件放在同一个文件夹下并运行程序文件，即可获得这两张人脸图片的相似度。同学们可根据实际情况准备更多的人脸图片，并修改程序中相应的参数，以获得更多人脸图片的比对结果，将比对情况及结果填入表4.4.1中，思考如何根据比对结果数据来判断两者是否为同一人。

表 4.4.1　人脸图片比对结果记录表

序号	人脸图片*m*的名称及说明	人脸图片*n*的名称及说明	比对结果（相似度）
1			
2			
3			

人脸识别及其应用的实现涉及的技术环节很多，如人脸检测、分析、比对、在线活体检测（判断有脸真实性是否为照片翻拍）、人脸库管理、身份验证等。其中的人脸检测与属性分析，可检测一图多脸图片上的人脸数量与部位并标记出边框，只要图片足够清晰，就可对图中的每张人脸分别进行分析，获得眼、口、鼻的轮廓等多个关键点定位，从而得出多种人脸属性，如性别、年龄范围、表情等信息。人脸检测分析技术已经较好地解决了大角度侧脸、遮挡、模糊、表情变化等问题，而且在各种实际环境中都有较好的适应性，已成功应用于智能相册分类、人脸美颜、互动营销等场景中。

※ 活动2　人脸库建设与管理

人脸识别应用的实现通常也包括人脸库维护管理，实际活动场景中的人脸捕捉与信息提取，在人脸库中检索、比对、获得识别结果等，在公共安全、各类考勤、身份认证等诸多领域有着广泛的应用。

活动场景中的人脸捕捉与信息提取一般是通过从摄像头视频流中抓帧并开展智能分析来实现的。人脸库维护则是把已知人员的信息、头像图片以及从头像图片中提取的特征保存到数据库中，并对这些信息进行分组管理。当需要对某个未知身份的人脸图片进行识别时，就可以从该人脸图片中提取特征并将其与数据库中的记录进行检索比对，按照事先设好的阈值筛选出匹配度最高的人员信息与照片序列，从而实现对该图片的人脸识别。因此，人脸库维护是人脸识别中一个非常重要的环节。

百度AI开放平台提供了人脸库的管理服务，包括人脸注册、人脸更新与删除、用户组管理等功能。通过调用相应API便可方便地创建、维护属于自己的人脸库。本活动中，我们将尝试把已知人员的信息与照片上传至人脸库，并开展相应的管理维护。

请事先准备好几张已知人员照片，如user1.jpg、user2.jpg、user3.jpg等，尝试用程序的方式开展人脸库的管理维护。

在Python的IDLE窗口中编写程序，在所属人脸库中创建用户组并导入用户照片，部分参考代码如下。

```
#初始化百度人脸识别实例
client = AipFace(app_id, api_key, secret_key)
client.groupAdd('users')                    #在所属人脸库中创建用户组users
#在users组中添加两个用户，参数：图片的base64值，图片类型，组id，用户id
user1 = base64.b64encode(open('user1.jpg', 'rb').read()).decode('utf-8')
user2 = base64.b64encode(open('user2.jpg', 'rb').read()).decode('utf-8')
client.addUser(user1, 'BASE64', 'users', 'ykm')
client.addUser(user2, 'BASE64', 'users', 'myd')
```

你也可以在百度AI开放平台中直接创建并管理你的人脸库，包括用户组的创建与删除和人脸照片的上传与删除。当然你也可以参考上述代码和相关技术文档，以程序方式管理你的人脸库用户组和用户照片。

※ 活动3　人脸搜索与识别

百度AI开放平台的人脸搜索API接口提供了在你的人脸库中的指定用户组中检索比对上传图片中人脸的功能，并且可以返回匹配度从高到低的用户信息，我们可以结合实际情况给匹配度设定一个阈值，只要匹配度高于这个阈值就是该图片的识别用户了。借助这个接口，我们便可以实现指定图片的人脸识别。下面就让我们一起来尝试。

实际应用场景中，待识别的人脸图片一般从摄像头拍到的视频流中通过智能抓帧分析获取。方便起见，我们可以手动拍摄几张作为识别对象的人脸图片（如img1.jpg、img2.jpg，需含有与前面人脸库中相同人员的照片），有条件的同学也可以在老师的帮助下从摄像视频流中通过抓帧获得图片。

在Python的IDLE窗口中编辑程序，实现上传待识别人脸图片并返回识别结果，部分参考代码如下。

```
#在初始化百度AipFace实例client的基础上，通过以下代码实现人脸搜索
#将待识别图片转为base64编码
image = base64.b64encode(open('img1.jpg', 'rb').read()).
decode('utf-8')
#从指定的users组中查找并返回结果
res = client.search(image, "BASE64", 'users')
#从返回结果中提取最匹配的用户ID
user = res['result']['user_list'][0]['user_id']
#从返回结果中提取最匹配用户相似度
score = res['result']['user_list'][0]['score']
print('在该用户组中匹配度最高的为用户 %s，匹配度为 %.1f' % (user, score))
```

请通过修改代码中的“img1.jpg”参数更换待识别图片，重新运行程序，将识别结果填入表4.4.2中。

表 4.4.2　人脸识别结果记录表

序号	人脸图片及名称	返回用户信息及匹配度	结果评判（正确与否）
1			
2			
3			

人脸识别的一般过程与原理

人脸识别系统主要包括四个组成部分，分别为人脸图像采集及检测、人脸图像预处理、人脸图像特征提取以及人脸图像匹配与识别。

人脸图像采集

不同的人脸图像都能通过照相机或摄像机采集下来，静态图像、动态图像、不同的位置、不同的表情等都可以被很好地采集。当用户在采集设备的拍摄范围内时，采集设备会自动检测搜索并拍摄用户的人脸图像。

人脸图像检测

人脸图像检测属于整个人脸识别流程的预处理环节，即在图像中准确标出人脸的位置和大小。人脸图像中包含的模式特征十分丰富，如直方图特征、颜色特征、模板特征、结构特征及Haar-like特征（一种用于人脸特征描述的模板，它反映了图像的灰度变化情况）等。人脸检测就是把这其中有用的信息挑出来，并判断这些特征用于检测该人脸是否足够有效。

人脸图像预处理

人脸的图像预处理是基于人脸图像检测结果，对图像进行处理并最终服务于特征提取的过程。由于受到各种条件的限制和随机干扰，系统获取的原始图像往往不能直接使用，必须对它进行灰度校正、噪声过滤等图像预处理。对于人脸图像而言，其预处理过程主要包括人脸图像的光线补偿、灰度变换、直方图均衡化、归一化、几何校正、滤波以及锐化等。

人脸图像特征提取

人脸图像特征提取，也称人脸表征，它是对人脸进行特征建模的过程。人脸图像特征提取方法一般分为两大类：一种是基于知识的表征方法，另外一种是基于代数特征或统计学习的表征方法。人脸识别系统可使用的特征通常分为视觉特征、像素统计特征、人脸图像变换系数特征、人脸图像代数特征等，人脸图像特征提取就是针对这些特征或部分特征进行的。

人脸图像匹配与识别

人脸识别就是将待识别的人脸特征与已得到的人脸特征模板进行比较，根据相似程度对人脸的身份信息进行判断。即将提取的人脸图像的特征数据与数据库中存储的特征模板进行搜索匹配，通过设定一个阈值，当相似度超过这一阈值时，则输出匹配得到的结果。这一过程又分为两类：一类是确认，是一对一进行图像比对的过程；另一类是辨认，是一对多进行图像匹配比对的过程。

4.5　语音识别及其应用

近年来，语音识别技术取得显著进步，开始从实验室走向市场，包括语音拨号、语音导航、设备控制、语音文档检索、简单的听写数据录入等诸多方面。语音识别技术与其他自然语言处理技术如机器翻译相结合，可以构建出更加复杂的应用，如从一种语言的语音到另一种语言的语音的翻译等。语音识别技术的发展趋势已经十分明显，很快将广泛应用于工业、家电、通信、汽车电子、医疗、家庭服务等各个领域。本节我们将借助人工智能开放平台开展语音识别技术的应用创新，实现将语音识别为文字及将文字转化为声音的人工智能应用。

学习目标

- ★ 掌握语音识别的方法。
- ★ 掌握语音合成的方法。
- ★ 体验简单的人机对话过程。

任务　实现语音识别并探讨其可能的创新应用

※ 活动1　实现简单的语音识别

语音识别是模式识别的一个分支，又从属于信号处理科学领域，同时它又与语音学、语言学、数理统计及神经生物学等学科有密切的关系。通过前面的活动，我们成功配置了人工智能开放平台API运行环境，接下来将尝试实现语音识别应用，测试语音识别效果。

请分组开展协作学习，完成以下活动，填写表4.5.1，然后以小组为单位在班级中进行分享。

表4.5.1 语音识别过程记录表

朗读次数	朗读的内容	识别结果
第1次		
第2次		
第3次		

第一步，各组分别调试好麦克风与计算机的连接，准备好三段60秒以内、不同内容的文字。

第二步，安装处理声音所需的Python运行库。

在Windows命令行窗口下运行以下代码，安装所需的Python程序包。

```
pip install SpeechRecognition
pip install PyAudio
pip install playsound
```

其中，SpeechRecognition是一套基于Python支持语音识别的包；PyAudio包支持录音、播放和生成wav文件等功能的实现，为SpeechRecognition所依赖；playsound包用于播放mp3等音频文件。

第三步，编写程序代码。

（1）启动Python IDLE，编写代码以module.py为文件名保存，部分参考代码如下。

```
class module():
#此处省略部分代码，详细内容参见教科书配套资源
  def listen(self):
    with sr.Microphone() as source:   #获取麦克风输入并处理环境噪声
      self.recognizer.adjust_for_ambient_noise(source)
      audio = self.recognizer.listen(source)  #听取麦克风输入数据
      return audio
  def recognize(self, audio):                 #将获取的声音识别为文字
    wav_data = audio.get_wav_data(16000)  #获取wav格式数据
    #上传识别并返回结果
    ret = self.bd_aip.asr(wav_data, 'wav', 16000, {'dev_pid': 1536})
    if ret['err_no'] == 0:                    #处理百度语音的返回结果
      return ret['result'][0]                 #返回中文识别结果
```

（2）在Python IDLE中新建run.py程序并保存，部分参考代码如下。

```
from module import Module
#此处省略部分代码，详细内容参见教科书配套资源
module = Module(app_id, api_key, secret_key)    #创建语音识别实例
while True:
  audio = module.listen()                       #获取麦克风输入数据
  result = module.recognize(audio)              #发送并识别成文字
  print(result)                                 #输出识别结果
```

第四步，将module.py和run.py置于同一个目录下，运行run.py文件，对着麦克风朗读准备好的文字，获得语音识别后的文本。

语音识别的基本原理

语音识别是以语音为处理对象，通过语音信号处理和模式识别让机器自动识别和理解人类口述的语言。语音识别技术就是让机器通过识别和理解过程把语音信号转变为相应的文本或命令的技术。语音识别是一门涉及面很广的交叉学科，它与声学、语音学、语言学、信息理论、模式识别理论以及神经生物学等都有非常密切的关系。语音识别技术正逐步成为计算机信息处理技术中具有重要发展前景的技术，语音技术的应用已经形成了一个非常有潜力的市场。

语音识别系统本质上是一种模式识别系统，包括特征提取、模式匹配、模型库等三个基本单元，它的基本结构如图4.5.1所示。

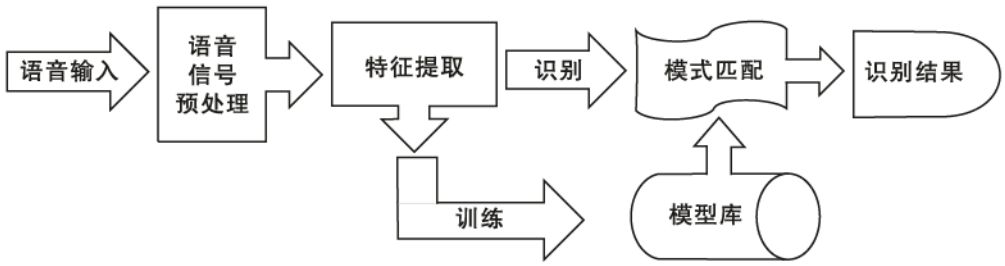


图 4.5.1　语音识别系统的基本结构

通过“模/数转换”（Analogue-to-Digital Conversion，ADC），将模拟信号的语音变换成数字化的电信号后，加在语音识别系统的输入端，首先经过预处理，再根据人的语音特点建立语音模型，对输入的语音信号进行分析，并抽取所需的特征，在此基础上建立语音识别所需的模板。在识别过程中要将计算机中存放的语音模板与输入的语音信号的特征进行比较，根据一定的搜索和匹配策略，找出一系列最优的与输入语音匹配的模板，然后根据此模板的定义，通过自动查表给出识别结果。显然，结果的优劣与特征的选择、语音模型的好坏、模板是否准确都有

直接的关系。

语音识别系统整体上包括两大部分，即训练和识别。训练通常是离线完成的，对预先收集好的海量语音进行信号处理和知识挖掘，获取语音识别系统所需要的“声学模型”和“语言模型”；而识别过程通常是在线完成的，对用户语音进行实时自动识别并给出实时反馈。识别过程通常又可以分为“前端”和“后端”两大模块：“前端”模块的主要作用是进行端点检测（去除多余的静音和非说话声）、降噪、特征提取等；“后端”模块的作用是利用训练好的“声学模型”和“语言模型”对用户说话的特征向量进行统计模式识别（又称“解码”），得到其包含的文字信息。此外，后端模块还存在一个“自适应”的反馈模块，可以对用户的语音进行自学习，从而对“声学模型”和“语音模型”进行必要的“校正”，进一步提高系统识别的准确率。

※ 活动2　实现简单的语音合成

在本活动中，我们要用不同的文本测试将文字合成为语音的效果。

第一步，编写程序代码。

在Python IDLE编辑页面中输入以下代码并保存为run.py文件。

```
from aip import AipSpeech
from playsound import playsound          #用于播放mp3音频
#通过secret_key授权码创建语音识别对象
client = AipSpeech(app_id, api_key, secret_key)
text = '轻轻的我走了,正如我轻轻的来.'    #填入要合成为语音的文字
result = client.synthesis(text)          #返回语音合成结果
with open('audio.mp3', 'wb') as f:       #将合成结果写入audio.mp3文件
  f.write(result)
playsound('audio.mp3')                   #播放audio.mp3音频文件
```

第二步，运行程序。

（1）运行程序，获得合成后的语音。

（2）修改程序中需要合成为语音的文本后再运行程序进行转换，获得合成后的语音。

（3）再次修改程序中需要合成为语音的文本后运行程序进行转换，获得合成后的语音。测试不同文本合成为语音的准确率。

上述活动完成后，请填写表4.5.2。

表 4.5.2　语音合成过程记录表

次数	文字内容	合成为语音的结果
第一次合成		
第二次合成		
第三次合成		

LD332X系列语音识别芯片的工作原理

语音识别属于数字信号处理的研究领域，其算法初期是依靠计算机、数字信号处理器等来实现的，但随着微电子学和集成电路技术的新进展，近年来不断有语音识别集成电路芯片出现，在提高实时识别效率方面有不俗的表现。其中，ICRoute的LD3320就是一款性价比较高的芯片。

LD3320提供的语音识别技术，是基于“关键词语列表”的识别技术，即ASR（Automatic Speech Recognition）技术。语音识别芯片的工作流程是：把通过MIC输入的声音进行频谱分析—提取语音特征—和关键词语列表中的关键词语进行对比匹配—找出得分最高的关键词语作为识别结果输出，如图4.5.2所示。

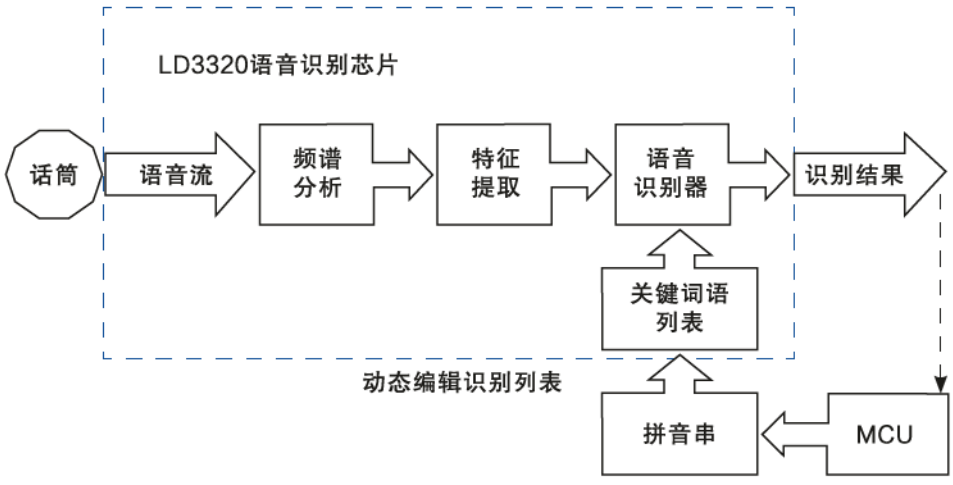


图 4.5.2　LD332X系列语音识别芯片工作原理

语音识别芯片能在两种情况下给出识别结果。

（1）外部送入预定时间的语音数据（比如5秒的语音数据），芯片对这些语音数据进行运算分析后，给出识别结果。

（2）外部送入语音数据流，语音识别芯片通过端点检测VAD（Voice Activity Detection）检测出用户停顿的间隙，来判断用户是否已

经说完一句话，把用户开始说话到停止说话之间的语音数据进行运算分析后，给出识别结果。

对于第一种情况，可以理解为设定了一段定时录音（比如为5秒），芯片在5秒后会停止把声音送入识别引擎，并且根据已送入引擎的语音数据计算出识别结果。

对于第二种情况，需要了解VAD的工作原理。

VAD技术是在一段语音数据流中，判断哪个时间点是人声的开始，哪个时间点是人声的结束。判断的依据是，如果在背景声音的基础上出现了语音发音，则视为声音的开始。之后，如果检测到持续一段时间的背景音（比如600毫秒），则视为人声说话结束。通过VAD判断出人声说话的区域后，语音识别芯片会识别处理这期间的声音数据，计算出识别结果。

与人不同，语音识别算法无法“主动”地判断出什么时候“应该”给出一个识别结果。这是因为在计算过程中的任何时刻，语音识别器都会对已送入识别芯片的声音数据进行分析，并根据匹配程度为识别列表中的关键词语打分。但是，由于识别算法不知道用户后面是否还继续说话，不知道该句话是否结束，所以无法“主动”地判断并给出结果。比如，识别列表中有两个关键词语分别是“于燕”和“于燕芳”，当用户说到“燕”这个音节时，在识别芯片内部是“于燕”的得分最高，但此时识别芯片还不能给出识别结果；随着用户说出“芳”音节后，“于燕芳”则成为得分最高的词。此时，用户停止说话，或者是定时录音已结束，使得识别芯片能判断出用户已经停止说话了，这样才能给出识别结果是“于燕芳”。

※ 活动3　尝试人机对话

在本活动中，我们可以尝试用语音来指挥计算机工作。例如，我们对计算机说“打开记事本”“打开画图”等，计算机便能准确地执行指令。

第一步，编写程序代码。

在Python IDLE编辑页面中输入代码并保存为talk.py文件：

```
import os
from module import Module
#通过secret_key等参数创建实例对象
module = Module(app_id, api_key, secret_key)
while True:
```

```
audio = module.listen()
result = module.recognize(audio)
print(result)
if result == '打开记事本':
  os.system('notepad')
elif result == '打开画图':
  os.system('mspaint')
```

第二步，运行程序。

运行talk.py，用语音对计算机发出指令，测试计算机执行语音指令的准确程度。

上述活动完成后，请填写表4.5.3。

表 4.5.3　语音识别过程记录表

对话内容	执行情况	备注
打开记事本		
打开画图		
重启计算机		

常见的人工智能音箱

智能音箱是音箱与人工智能相结合的产物，是用语音进行交互的一个智能工具。一个典型的智能音箱在硬件上应该具备语音的输入、输出模块、Wi-Fi模块、语音算法本地处理单元、音效单元、充电及其他外设接口等，在软件上应该具备语音算法运行程序、服务API接口程序、手机端APP等。智能音箱一般具备语音交互体验、有声资源播放、智能家居控制、生活O2O服务和生活小工具等五大功能。一个典型的智能音箱如图4.5.3所示。

图 4.5.3　一个典型的智能音箱

目前市场上智能音箱的品种非常多，具有代表性的智能音箱有以下几种。

亚马逊Echo

亚马逊推出搭载语音平台Alexa的智能音箱Echo，首创智能音箱先河。Echo最大的亮点是将智能语音交互技术植入传统音箱中，从而赋予音箱人工智能的属性。用户能直接通过语音操控它，而且还可以通过它来控制智能家居。Alexa作为语音助手，可以像朋友一样与用户交流，同时还能实现播放音乐与新闻、网购下单、叫车、订外卖等功能。

小度智能音箱

小度智能音箱搭载了百度对话式人工智能操作系统DuerOS，拥有超过1000万小时的海量有声内容，400多项生活常用技能。它能通过语音搜索歌手名称、歌曲名称、歌词；支持天气查询、限号记录查询、数学题计算、股市查询、货币汇率计算，提供备忘录、倒计时、闹钟等实用功能；可接受通过语音控制常用的家居产品，如灯、空调、空气净化器、热水器、窗帘等。

小米AI音箱

小米公司把“小爱同学”作为小米智能音箱的唤醒词。小米智能音箱可以和米家APP中绑定的部分小米智能家居产品进行捆绑，授权之后，用户可以通过音箱的语音交互对它们进行控制。小米AI音箱支持语音交互，支持在线音乐、网络电台、有声读物、广播电台等，提供新闻、天气、闹钟、倒计时、备忘、提醒、时间、汇率、股票、限行、算数、查找手机、百科/问答、闲聊、笑话、菜谱、翻译等服务。

单元学习评价

本单元我们学习了拍照识花、简单中文文本挖掘、手写体字符识别、人脸识别和语音识别与合成等人工智能应用的相关知识，以及借助人工智能应用框架和开放平台API开发简单人工智能应用的一般方法。结合本单元的学习体验，请回答以下问题。

1. 2017年被列为国内首批四大人工智能开放平台的是哪些？它们各有什么特色与侧重点？常见的人工智能应用框架有哪些？

2. 文本挖掘是指从大量文本数据中抽取_____、_____和_____的信息或知识的过程。它可广泛应用于_____和_____、文本聚类、_____、Web挖掘、信息抽取和_____等方面。

3. Sklearn是一个开源的通用机器学习工具包，它涵盖并实现了几乎所有主流_____算法，针对从数据预处理到训练模型的各个方面，包括_____、_____、_____和_____等，并提供了统一的调用接口。

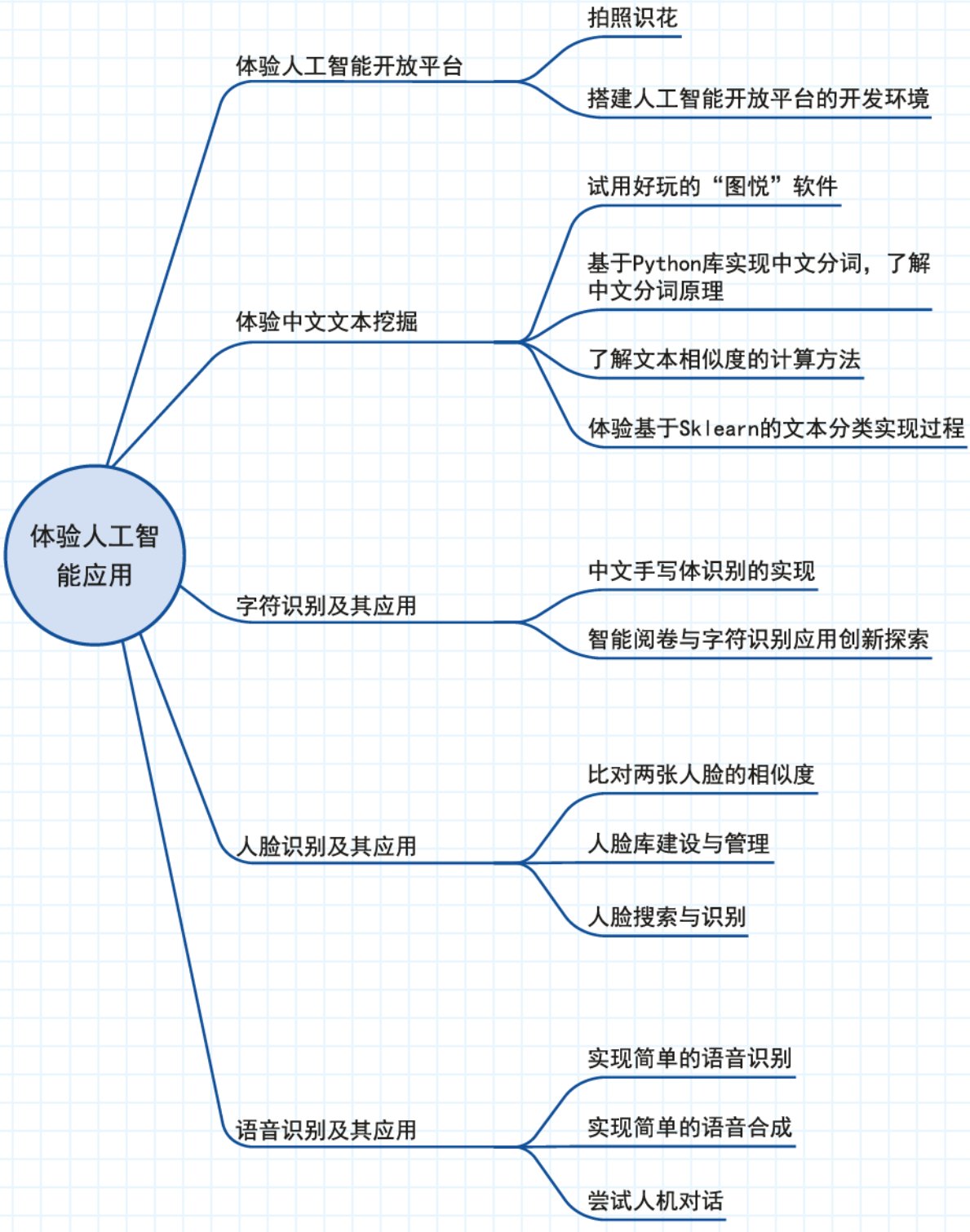
4. 字符识别OCR的全称是_____，是机器学习_____应用的一种，是指对文本资料的_____进行分析与识别处理，获取_____及_____的过程。在整个字符识别过程中，对识别准确率影响最大的技术瓶颈一般是_____和_____两个环节。

5. 人脸识别是基于人的_____进行身份识别的一种生物识别技术，也称为人像识别或面部识别。它包含借助摄像机等硬件直接_____或从视频流中_____，再在图像中_____人脸，进而对检测到的人脸进行_____等环节。人脸识别的技术层面又可依次分为_____、_____和_____。

6. 人脸识别系统主要包括_____、_____、_____和_____四个组成部分。

7. 语音识别是以_____为处理对象，通过_____处理和_____让机器自动识别和理解人类口述的语言。语音识别技术就是让机器通过识别和理解过程把_____转变为相应的_____或_____的技术。语音识别系统本质上是一种模式识别系统，包括_____、_____和_____三个基本单元。

单元学习总结


体验人工智能应用
体验人工智能开放平台
拍照识花
搭建人工智能开放平台的开发环境
体验中文文本挖掘
试用好玩的“图悦”软件
基于Python库实现中文分词，了解中文分词原理
了解文本相似度的计算方法
体验基于Sklearn的文本分类实现过程
字符识别及其应用
中文手写体识别的实现
智能阅卷与字符识别应用创新探索
人脸识别及其应用
比对两张人脸的相似度
人脸库建设与管理
人脸搜索与识别
语音识别及其应用
实现简单的语音识别
实现简单的语音合成
尝试人机对话

第 5 单元　人工智能的辩证思考

经过60多年的演进，人工智能的发展方兴未艾。2017年7月国务院印发的《新一代人工智能发展规划》中指出，人工智能正呈现出深度学习、跨界融合、人机协同、群智开放、自主操控等新特征。规划中还明确了我国新一代人工智能发展的战略目标：到2020年，人工智能总体技术和应用与世界先进水平同步，人工智能产业成为新的重要经济增长点，人工智能技术应用成为改善民生的新途径；到2025年，人工智能基础理论实现重大突破，部分技术与应用达到世界领先水平，人工智能成为带动我国产业升级和经济转型的主要动力，智能社会建设取得积极进展；到2030年，人工智能理论、技术与应用总体达到世界领先水平，成为世界主要人工智能创新中心。

毫无疑问，人工智能预示着美好的未来。然而，人工智能和其他事物一样，也有着其两面性，在造福人类的同时，也可能会被不法分子利用而危害他人。在一个充斥了人工智能产物的社会中，人工智能可以很好，也可以很坏，它与我们近在咫尺。如何掌控人工智能，让其充分发挥正向价值，服务于人类的美好生活，是一个值得思考的问题。

本单元中，我们将围绕“辩证地认识人工智能的巨大价值和潜在威胁”项目开展学习，通过智能系统的应用实践，结合数字化学习的体验，了解智能应用系统在给人类带来便利的同时所面临的伦理及安全挑战，辩证地认识人工智能这把“双刃剑”，思考如何掌控它，尽可能地避免负面影响，让它真正地为人类服务。

为了完成该项目，需要思考以下问题：智能应用系统所面临的伦理及安全挑战有哪些？人工智能的发展趋势是怎样的？有哪些潜在威胁？人工智能在未来人类社会有哪些应用价值？人工智能可能会带来哪些新的安全问题？人工智能在保护信息安全方面有哪些作用？如何看待人工智能对人类社会法律法规与伦理道德的挑战？为此，我们需要完成以下具体任务：

- 通过实践了解智能应用系统对人类的帮助
- 通过调研了解人工智能的巨大价值和潜在威胁
- 调研并讨论人工智能对法律法规与伦理道德的挑战
- 自觉维护和遵守人工智能社会化应用的规范与法规

5.1 智能系统的应用体验

从互联网到物联网，从智能家居到智能楼宇，从智能小区到智慧城市，人们穿梭在物物相联的城市楼宇之间，用语音指示导航系统规划最优的上班路线，用指纹打卡，刷脸通过楼宇门禁，用语音指示个人助手订火车票、查找最近的酒店……。这一切都离不开智能系统的支持，这些智能技术越来越多地充当着人们助手的角色，极大地方便了人们的工作、生活与娱乐。所有事物都具有两面性，我们在享受着智能应用系统的便利之时，也同时面临着智能应用系统所带来的问题与挑战。

学习目标

★ 体验并辩证地认识常见的智能应用系统。

★ 了解智能应用系统所面临的挑战及应对措施。

★ 增强安全防护意识和责任感。

当前常见的智能系统有很多，比如用声音验证的声纹锁、用指纹验证的指纹锁等。下面我们一起来体验这些智能系统的简单应用并了解智能系统可能会遇到的一些安全问题。

任务 通过实践体验常见智能系统的应用

※ 活动1 声纹锁的安装与使用

生物特征认证——声纹认证，是用每个人独一无二的生物特征来验证用户身份的技术，不容易被仿冒也不可能遗失。

用独有的声纹锁软件把计算机系统保护起来是比较安全的方法。安装好声纹锁软件之后，需要进行声纹训练与学习。用户需要重复朗读系统提示的文字，以达到有效的声纹录入。某声纹录入系统交互界

面如图 5.1.1 所示。声纹注册完成后，在下次进入系统时，就要进行声纹认证，只有声纹完全符合的用户才被允许进入系统。

图 5.1.1　声纹注册与设置

按表5.1.1所示的主题划分小组，在老师的指导下，使用移动终端（如手机或PAD）下载、安装一款声纹锁软件并打开使用，分组检测声纹锁的安全性能，讨论声纹锁是否有被破解的可能。

要求每个小组针对上述主题开展协作学习，并完成相关主题的知识获取、筛选与整理，然后以小组为单位在班级中进行分享与讨论。

上述活动完成后，请填写表5.1.1。

表 5.1.1　声纹锁使用活动记录表

序号	主题	结果记录	备注
1	声纹锁软件的种类		
2	声纹锁软件的注册		
3	声纹锁软件的使用		
4	声纹锁软件的设置		
讨论：声纹锁有可能被破解吗？			

※ 活动2　指纹锁的安装与使用

生物特征认证——指纹认证，是用每个人独一无二的指纹特征作为生物特征来验证用户身份的技术，同样不容易被仿冒也不可能遗失。

用独有的指纹锁软件把计算机系统保护起来是比较安全的方法。安装好指纹锁软件之后，需要进行指纹注册。用户需要重复按压指纹，以达到指纹正确录入。用户可以注册多个手指的指纹。一个指纹认证系统的注册界面如图5.1.2所示。指纹注册完成后，在下次进入系统时，就需要进行指纹认证，只有指纹完全符合的用户才被允许进入系统。

图 5.1.2　指纹注册与设置

按表5.1.2所示的主题划分小组，在老师的指导下，使用移动终端（如手机或PAD）下载、安装一款指纹锁软件并打开使用，分组检测指纹锁的安全性能，讨论指纹锁是否有被破解的可能。

要求每个小组针对上述主题开展协作学习，并完成相关主题的知识获取、筛选与整理，然后以小组为单位在班级中进行分享与讨论。

上述活动完成后，请填写表5.1.2。

表 5.1.2　指纹锁使用活动记录表

序号	主题	结果记录	备注
1	指纹锁软件的种类		
2	指纹锁软件的注册		
3	指纹锁软件的使用		
4	指纹锁软件的设置		
讨论：指纹锁有可能被破解吗？			

拓展练习

体验指纹加密U盘或移动硬盘的使用方法，以及指纹加密的安全性。请分成A、B两组，各自设置指纹密码，并尝试进入指纹加密的U盘或移动硬盘，查看其中的文件，记录成功进入的概率。然后交换设备，尝试进入对方的指纹加密U盘或移动硬盘，查看其中的文件，看能否完成任务，做好记录，并说明理由。

※ 活动3　变声软件的安装与使用

利用变声软件可以将真实的声音变成假声音，也可以将一种声音类型变成其他的类型，如将真实的男声变成女声，模仿十几岁女孩、30或60岁女人的声音效果等。变声软件还提供了很多功能，如可以直接将歌曲或语音文件进行声音效果转换，利用可爱等类型进行修饰；能录制变声，可以在变声软件里录制自己的声音文件并进行配置变声；能模仿变声，可以将不同的声音进行合成，从而模仿出某个特定的人的声音，或者创建出一个全新的声音组合，并保存在变声库中，供后续使用。为支持网络应用，变声软件往往会兼容各类社交软件、直播平台、游戏平台等各类平台。某变声软件的实例如图5.1.3所示。

图 5.1.3　电话变声器

按表5.1.3所示的主题划分小组，在老师的指导下，使用移动终端或个人计算机下载、安装一款变声软件并打开使用，分组检测变声软件的变声效果。

要求每个小组针对上述主题开展协作学习，并完成相关主题的知识获取、筛选与整理，然后以小组为单位在班级中进行分享与讨论。

上述活动完成后，请填写表5.1.3。

表 5.1.3　变声软件的安装与使用活动记录表

序号	主题	结果记录	备注
1	电话变声软件的安装		
2	电话变声软件的使用		
3	聊天变声软件的安装		
4	聊天变声软件的使用		
讨论：变声软件各有什么利弊？			

● 华丽琴鸟

琴鸟是澳大利亚的国鸟。它能歌善舞，歌声婉转动听，舞姿轻盈优美，不但能模仿各种鸟类的鸣叫声，还能学人间的各种声音，如汽车喇叭声、火车喷气声、斧头伐木声、修路碎石机声及领号人的喊叫声等。

国外有一家名为“华丽琴鸟”的公司，是首家利用一小段声音作为样本便可准确复制该人声音的公司。他们以蒙特利尔大学博士生们研发

的深度学习模型为基础，将一段语音中的个人特征压缩成一段独特的编码，将该编码输入算法后，不到半秒便可生成1000个句子。该算法不仅能合成语音，还能对声音进行控制，赋予其喜悦、愤怒、同情或紧张等情感色彩。该公司官网上以特朗普、奥巴马和希拉里的声音为例，演示了该技术效果的逼真程度。研发人员称，这一技术可得到广泛运用，如充当个人助理、用名人的声音阅读有声书或为残疾人合成“演讲”等，在动漫电影和视频游戏中也将有用武之地。

该公司的官网提供了测试的版本，用户只要朗读30条系统提供的句子并上传，系统就可以生成具有用户个人特征的声音库，当用户再输入文字的时候，系统就能以用户的声音来朗读、输出这些句子。该系统的界面如图5.1.4所示。

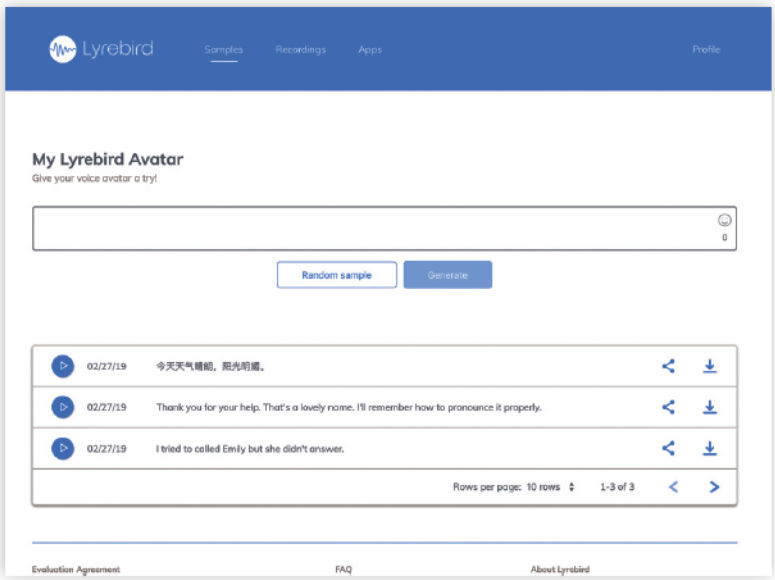


图 5.1.4　输出具有个人特征的声音

但是这样的技术也有可能引发严重的社会问题。研究人员在官网上写道：录音常被视作强有力的证据，许多国家的司法系统尤其看重这一点，而不法分子可能会利用我们发明的技术轻易操纵录音，从而破坏录音作为证据的可信度，该技术可能导致危险后果，如“通过窃取他人身份误导外交官或进行欺诈等”。该团队认为，等到该技术对公众开放之后，录音便不应再被视作验明正身的证据。

※ 活动4　了解机器学习在入侵检测系统中的应用

按表5.1.4所示的主题划分小组，了解机器学习在入侵检测系统中的应用等相关知识。在老师的指导下，设计出高效的搜索关键词；在老师的推荐下，选择较权威的官方网站，开展数字化学习。

要求每个小组针对上述主题开展协作学习，并完成相关主题的知

识获取、筛选与整理，然后以小组为单位在班级中进行分享与讨论。

上述活动完成后，请填写表5.1.4。

表 5.1.4　机器学习在入侵检测系统中的应用活动记录表

序号	主题	结果记录	备注
1	入侵检测系统的种类		
2	入侵检测系统的工作原理		
3	入侵检测系统中机器学习的内容		
4	入侵检测系统中机器学习的方法		
5	入侵检测系统中机器学习的原理		

入侵检测系统

入侵检测系统（Intrusion Detection System，IDS）是一种对网络传输进行即时监视，在发现可疑传输时发出警报或者采取主动反应措施的网络安全设备。入侵检测系统根据信息来源的不同和检测方法的差异分为几类：根据信息来源可分为基于主机的入侵检测系统和基于网络的入侵检测系统，根据检测方法又可分为异常入侵检测和误用入侵检测。

入侵检测系统包括四个组件：事件产生器（Event Generators），它从整个计算环境中获得事件，并向系统的其他部分提供此事件；事件分析器（Event Analyzers），它经过分析得到数据，并产生分析结果；响应单元（Response Units），它对分析结果做出反应，例如切断连接、改变文件属性等，也可以只是简单地报警；事件数据库（Event Databases），存放各种中间和最终数据，它可以是复杂的数据，也可以是简单的文本文件。

当前网络发展迅速，网络传输速率大大加快，给入侵检测系统的工作造成了很大负担，这也意味着入侵检测系统对攻击活动检测的可靠性不高。入侵检测系统在应对攻击时，也会抑制对其他传输的检测。同时由于模式识别技术的不完善，入侵检测系统的高虚警率也是它的一大问题。高虚警率会形成“狼来了”效应。为了开发更智能的入侵检测系统，需要将机器学习应用于入侵检测、木马检测、漏洞扫描等方面。

人工智能入侵检测系统

在互联网迅速发展的信息时代，信息安全问题日益严峻，并引起了人们的高度关注。如何将人工智能技术应用到信息安全领域中，已经被很多业界专家所重视。

基于机器学习的人工智能入侵检测系统逐渐进入了人们的视野。基于G-means算法机器学习的人工智能入侵检测、木马检测、漏洞扫描等方面的应用原理如图5.1.5所示。其中，网络数据包捕获模块用来实现监视和验证网络实时工作状态与流量；数据预处理模块从中提取能够代表数据特征的信息组成模式样本后再用于训练和检测；G-means机器学习模块是该系统的核心，通过该模块的训练，机器能够检测入侵。

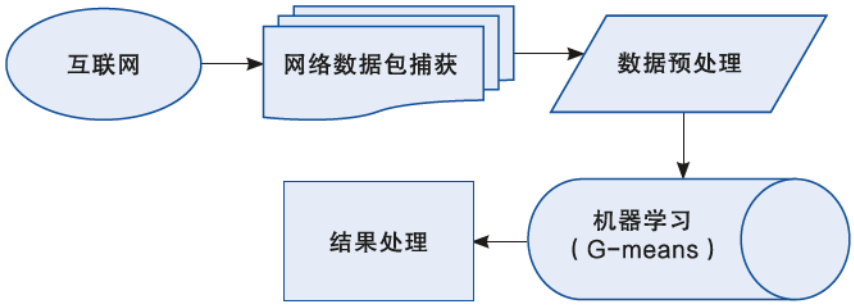


图5.1.5 基于G-means算法的网络入侵检测系统

机器学习的方法很多，在上述模块的训练中选用了G-means算法。在前期的训练阶段，将归一化的特征向量作为G-means的输入进行训练；训练完毕后，训练结果将作为后期实时在线的入侵检测部分的事件判定标准信息。

在图5.1.5中，结果处理模块对应入侵检测系统标准模型的事件分析器和响应单元两个部分，它是入侵检测系统中的主动武器。入侵检测系统会根据其机器学习模块在前期训练的结果信息，对接收网络数据包的信息进行在线实时分析和处理，判断是否是入侵行为或异常现象。若是，则将判断的结果转换为警告信息，然后根据警告信息做出反应，对事件发生采取措施，包括行为记录、原始数据特征记录、切断网络连接、终止用户进程、改变网络或系统存取属性等，也可以只是简单地报警，具体操作由安全管理人员实施。

机器学习在一线安全数据评估应用中取得重大进展，可使分析人员能够专注于对威胁进行更高级的调查，而不是执行战术性的数据处理。未来的可能情境是，入侵者会使用开源的人工智能和机器学习工具来找寻漏洞攻击企业或机构，而企业、机构和政府也同样需要借助人工智能技术的力量来打击网络犯罪，人工智能将使网络空间中的安全问题成为全新的战场。

5.2　人工智能的巨大价值和潜在威胁

在人工智能技术高速发展的今天，我们也听到了业界很多专家对人工智能的担忧。一直致力于探索宇宙奥秘的霍金，对人工智能做出的预言是：未来人工智能也许会是人类的终结者！微软公司创始人比尔·盖茨认为，人类必须关心不久的未来人工智能可能带来的威胁。特斯拉首席执行官马斯克公开表示了对人工智能未来的担忧，他表示人工智能是现今存在的最大威胁，研究人工智能如同召唤恶魔。人工智能安全吗？人工智能带来的生产生活方式发生了深刻的变革，给信息系统安全、法律法规和伦理道德带来了哪些挑战？现行信息系统安全防范及法治体系又该如何调整和应对？

学习目标

★ 了解人工智能的巨大价值。

★ 了解人工智能的潜在威胁。

★ 增强人工智能应用情况下的遵规守法意识。

任务一　通过调研了解人工智能的巨大价值和潜在威胁

※ 活动1　了解人工智能对人类社会的巨大价值

按表5.2.1所示的主题划分小组，了解人工智能应用价值的相关知识。在老师的指导下，设计出高效的搜索关键词，如人工智能的应用价值；在老师的推荐下，选择权威性较高的电子书籍，如吴军的《智能时代——大数据与智能革命重新定义未来》，开展数字化学习。

要求每个小组针对上述主题开展协作学习，并完成相关主题的知识获取、筛选与整理，然后以小组为单位在班级中进行分享与讨论。

上述活动完成后，请填写表5.2.1。

表5.2.1 人工智能的应用价值调研活动记录表

序号	人工智能的应用场景	应用案例	成果与价值及未来发展
1	在农业中的应用	以色列农业：利用人工智能精准滴灌技术，人们在水资源严重匮乏的贫瘠土地上创造了农业出口大国的奇迹，以色列获得“欧洲的厨房”之称	欧洲第二大花卉供应国；农业出口占40%，平均一个国民贡献了世界上1.7个人的食物； 农业+人工智能=新型农业
2	在制造业中的应用		
3	在律师业中的应用		
4	在零售业中的应用		

人工智能的未来应用与价值

《新一代人工智能发展白皮书》指出，2017年全球人工智能核心产业规模已超过370亿美元，其中我国的人工智能核心产业规模已达到56亿美元左右。在下一阶段，得益于技术的持续进步和商业模式的不断完善，全球人工智能市场需求将进一步快速释放，预计带动2020年全球人工智能核心产业规模超过1300亿美元，年均增速达到60%；其中，我国人工智能核心产业规模将超过220亿美元，年均增速接近65%。

未来的医疗：在医疗领域，具有人工智能的计算机不仅能帮助诊断疾病，承担放射科医生的工作，还可以进行手术。今天，世界上最有代表性的做手术的机器人就是达·芬奇手术系统。与医生相比，计算机在诊断和做手术等方面具有人类不可比拟的优势：它们准确性高，且随着数据量（病例）的增加，其准确性提高很快；它们稳定性好，不会像人那样受情绪的影响。IBM开发的沃特森（Watson）智能系统可以理解自然语言，分析各种数据和医学影像，帮助诊断疾病和管理医疗信息。在一些医学领域，比如肿瘤科，它能够非常准确地给医生提供诊断的建议和帮助。随着医疗数据的高速增长，加上计算机学习能力的增强，这一类系统的进步会非常快。可以预见，在不久的将来，计算机在一些疾病的诊断方面会超过人。在未来的医疗中，将会出现更多性能更好的人工智能诊断与手术系统来造福人类。

达·芬奇手术系统是由总部设于美国加利福尼亚州桑尼威尔的Intuitive Surgical公司开发完成的。它的机械手臂的灵活性远远超过人，而且带有摄像机，可以进入人体内手术，因此不仅手术的创口非常小，而且能够实施一些人类医生很难完成的手术。目前全世界共装配了3000多台达·芬奇手术机器人，完成了300万例手术。

2012 年，谷歌科学比赛的第一名授予了一位来自美国威斯康星州的高中生，她通过对 760 万个乳腺癌患者的样本数据的机器学习，设计了一种确定乳腺癌癌细胞位置的算法，位置预测准确率高达 96%，超过了目前专科医生的水平。这位年轻学生采用的图像处理和机器学习算法都不复杂，她的成功完全得益于大数据，没有哪个大夫一生能够见识 760 万个病例。

未来的金融：在金融领域，智能个人身份识别将用于解决金融安全隐患，智能高频交易将用于提高金融决策效率，智能投资顾问将帮助金融机构开拓用户。

未来的交通：在交通领域，人工智能将应用于无人驾驶、智能汽车、交通规划等场合，用于解决目前交通行业普遍存在的驾驶感受差、道路严重拥堵等问题。

未来的教育：在教育领域，K-12线上教育以及大学配套设施等人工智能应用被学校和学生广泛使用，机器人成为广受欢迎的教育设备，智能辅导系统（ITS）也成为与科学、数学、语言学以及其他学科相匹配的互动导师。

未来的安全：在公共安全领域，人脸识别将广泛应用于安防监控，无人机、预测警务技术可以应用于反恐、维护公共治安等场景，用以解决公共安全隐患。

拓展练习

按表5.2.2所示的主题划分小组，结合已掌握的有关人工智能的知识，畅谈人工智能的美好未来。

要求每个小组针对上述主题开展协作学习，并完成相关主题的知识获取、筛选与整理，然后以小组为单位在班级中进行分享与讨论。

上述活动完成后，请填写表5.2.2。

表 5.2.2　畅想人工智能的美好未来活动记录表

序号	畅想人工智能的美好未来	案例畅想	成果与价值畅想
1	Google能否战胜死神	使用大数据找到衰老基因并修复基因，从而延长人类的寿命	李文森博士和Google共同创建的大数据医疗保健公司Calico帮助人类战胜死神
2			
3			
4			

※ 活动2　了解人工智能的未来发展和潜在威胁

按表5.2.3 所示的主题划分小组，在老师的指导下，设计出高效的搜索关键词，如弱人工智能、超人工智能；在老师的推荐下，选择权威性较高的官方网站，开展数字化学习，了解人工智能的未来发展阶段及其特征等相关知识。

要求每个小组针对上述主题开展协作学习，并完成相关主题的知识获取、筛选与整理，然后以小组为单位在班级中进行分享与讨论。

上述活动完成后，请填写表 5.2.3。

表 5.2.3　人工智能的未来发展阶段及其特征活动记录表

序号	分类	特征	案例
1	弱人工智能	特定领域，感知与记忆存储，如图像识别、语音识别	汽车的防抱死系统和控制喷油系统；手机的导航软件、天气预报、语音助手；网购时的商品推荐；棋类游戏等
2	强人工智能		
3	超人工智能		

人工智能的未来发展

阿里云研究中心的报告《人工智能未来制胜之道》指出，从人工智能的技术突破和应用价值两个维度分析，未来人工智能将会出现以下三个阶段，如图5.2.1所示。

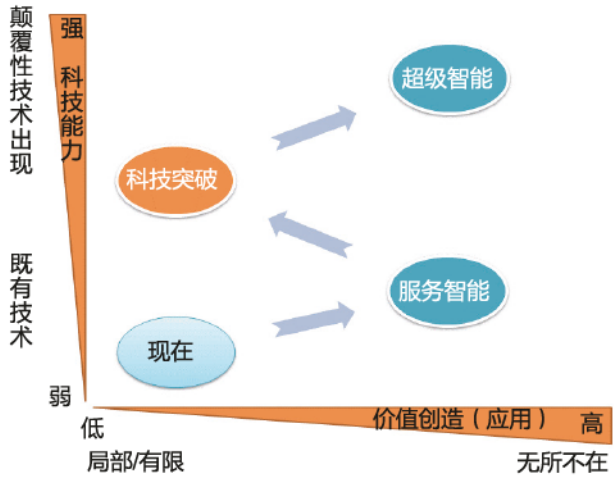


图 5.2.1　人工智能的未来发展阶段

第一阶段：当前，人工智能以服务智能为主，在人工智能既有技术的基础上，科技得以持续地进步，机器始终作为人的辅助；在应用层

面，人工智能拓展、整合多个行业应用，丰富使用场景。随着数据和场景的增加，人工智能创造的价值呈现指数式增长。

第二阶段：中长期将出现显著的科技突破。人工智能技术将取得显著突破，如自然语言处理技术可以即时理解人类对话，甚至可能预测出“潜台词”。在技术创新的领域，现有的人工智能技术应用向纵深拓展，但价值创造只限制在技术取得突破的领域。

第三阶段：长期可能出现超级智能。人工智能的技术取得显著突破，应用范围显著拓宽，人机完全共融，人工智能全面超越人类，无所不在，且颠覆各个行业和领域，创造极高的价值。

目前人工智能还只停留在“专有人工智能”阶段，主要是完成具体任务。现阶段，人工智能将逐渐向“通用人工智能”过渡，应用于完成复杂任务，如“识别医学图像，并快速诊断疾病（不限于肿瘤）”。中长期，随着技术显著突破，人工智能将逐步发展为“抽象人工智能”，在基础科技取得重大突破后，人工智能可以理解用户情感，从而改变用户行为，如“说服慢性病患者坚持按医嘱服药并在患病后改变生活习惯”。在遥远的将来，人工智能可能演变为超级人工智能，全面超越人类，通过技术突破和广泛的应用，预测并预先改变用户的行为，如“预先说服用户改变不良生活习惯，预防慢性病”。

未来，人工智能的应用将更有深度，并产生新的社会、商业和个人生活模式。人工智能的发展也呈现融合趋势：实现“感知/交互—正确理解—自主决策—自我学习”的实时循环；数据传输速度实现质的飞跃，云端将无缝融合；介入式芯片等新的硬件形式将出现，甚至实现人机共融。

人工智能的快速发展在获得关注与期待的同时，也在不断面对质疑与忧虑之声。霍金在GMIC大会（全球移动互联网大会）上提出了人工智能威胁论，即人工智能的崛起可能是人类文明的终结。霍金对人工智能的潜在威胁提出了警告：“超智能人工智能的出现将成为人类有史以来最好或最坏的事情。人工智能的真正风险不是恶意，而是能力。一个超级聪明的人工智能将非常善于实现其目标，如果这些目标与我们的目标不一致，我们就会陷入困境。可怕的现实是我们还不知道人工智能对世界是好还是灾难。”牛津大学人类未来研究院院长尼克·波斯特洛姆在《超级智能——路线图、危险性与应对策略》一书中亦专门讨论了人工智能的危险性问题。

为了缓解人工智能的潜在威胁，一份由14个权威机构著名专家联合撰写的报告《人工智能的恶意用途：预测、预防和缓解》给出了5点

建议：

（1）人工智能和机器学习研究人员应该承认其研究成果是双刃剑；

（2）政策制定者应该与技术人员密切合作，调查、预防和缓解人工智能可能的恶意使用方式；

（3）应该向计算机安全等其他高风险技术领域学习一些方法，将其应用于人工智能领域；

（4）应该在这些领域优先形成规范和道德框架；

（5）讨论这些挑战时所涵盖的利益相关者和专家范围应该扩大。

脑科学与人工智能

现在正处于人工智能发展的关键阶段，类脑智能、类脑芯片、脑机结合、深度学习算法、大规模神经网络建模等如雨后春笋般迅速发展。但人类对脑的认识依然很肤浅，人工智能向人脑的学习还不够，这是人工智能技术一直没有实现根本性突破的主要原因。只有清晰地了解大脑是如何运作的，才有可能设计出一台机器，做人类大脑能做的事情。

科学家们希望在人脑研究中取得重大进展，更深入地解析人脑工作的模式，为人工智能发展的革命性突破提供基础。脑科学被视为理解自然现象和人类自身的“终极疆域”，也为发展类脑计算系统和器件、突破传统计算机架构的束缚提供重要依据。

人工智能自诞生之初便奠定了其模拟、延伸、扩展人类智能的宏伟目标。近年来，人工智能研究的许多重要进展反映了一个趋势：来自脑科学的启发，即使是局部的借鉴也能够有效地提升现有人工智能模型与系统的智能水平。但要真正实现逼近乃至超越人类水平的人工智能，还需要对脑信息处理机制进行更为深入的研究和借鉴。类脑智能研究的目标就是通过借鉴脑神经结构及信息处理机制，实现机制类脑、行为类人的下一代人工智能系统。

类脑智能是以计算建模为手段，受脑神经和人类认知行为机制启发，并通过软硬件协同实现的机器智能。类脑智能研究未来在仿人运动模型、类人神经运动控制、人机协同的智能机器人控制等方面有望取得重大突破。

Google“人脑模拟器”通过将1.6万片电脑处理器（CPU）连接起来，创造了一个拥有 10 亿多条神经元连接的神经网络。国内首个100亿规模“神经元”人脑模拟器Westwell Brain也已经诞生。从模拟“神经元”的数量级上看，Westwell Brain模拟的“神经元”数量为Google“人脑模拟器”的10倍，是目前世界上在最小体积单位上实现最多神经元数

量的人脑模拟器。Westwell Brain与Google“人脑模拟器”最本质的区别在于，Westwell Brain已摆脱了程序存储控制结构，使用的是电路模拟“神经元”的方法，而Google“人脑模拟器”仍使用了1.6万片电脑处理器连接。

拓展练习

调查：强人工智能和超人工智能会在哪一天到来？

每一个弱人工智能的创新，都是给强人工智能和超人工智能大厦添砖加瓦。有专家认为，现在的弱人工智能，相当于地球早期软泥中的氨基酸——没有动静的物质，但突然之间就组成了生命。强人工智能和超人工智能会在哪一天到来？

以“未来的超人工智能有多可怕”“如果超人工智能出现，人类任何试图控制它的行为都是可笑的”“超人工智能不会出现”等关键词句上网搜索，开展数字化学习，将学习结果填入表5.2.4。

按表5.2.4所示分成“不会实现”和“会实现”两组，将学习所得的知识及自己的见解制作成演示文稿，然后进行分享、辩论与交流。

表 5.2.4　强人工智能和超人工智能到来时间点调查活动记录表

强人工智能实现的时间	
_____年	□不会实现
超人工智能实现的时间	
_____年	□不会实现

任务二　调研并讨论人工智能对法律法规与伦理道德的挑战

※ 活动1　遵守人工智能社会化应用的法律法规

按表5.2.5所示的主题划分小组，在老师的指导下，设计出高效的搜索关键词，如人工智能的法律三问；在老师的推荐下，选择权威性

较高的官方网站，开展数字化学习。通过数字化学习，了解人工智能的法律法规问题。

要求每个小组针对上述主题开展协作学习，并完成相关主题的知识获取、筛选与整理，然后以小组为单位在班级中进行分享与讨论。

上述活动完成后，请填写表5.2.5。

表5.2.5 人工智能的法律法规问题活动记录表

序号	问题	案例	答案
1	人工智能生成物是否具有知识产权	2017年5月，“微软小冰”创作的诗集《阳光失了玻璃窗》出版	人工智能生成物具备知识产权作品的某些属性
2	人工智能可以替代司法者吗		
3	人工智能侵权责任如何认定		

人工智能侵权责任如何认定？

2016年11月，在深圳举办的第十八届中国国际高新技术成果交易会上，一台名为“小胖”的机器人突然发生故障，在没有指令的情况下，自行砸坏了部分展台，并导致一人受伤。人工智能应用引发的侵权责任认定问题，是对现行侵权法律制度提出的又一个新的挑战。

从现行法律上看，侵权责任主体只能是民事主体，人工智能本身还难以成为新的侵权责任主体。即便如此，人工智能侵权责任的认定也面临诸多现实难题。侵权发生后，谁是人工智能的所有者，就应当由谁负责，在法律上似乎并不存在争议。然而人工智能的具体行为受程序控制，发生侵权时，到底是由所有者还是软件研发者担责，值得商榷。

与之类似的，当无人驾驶汽车对他人造成损害或侵权时，是由驾驶人、机动车所有人担责，还是由汽车制造商、自动驾驶技术开发者担责？法律是否有必要为无人驾驶汽车制定专门的侵权责任规则？这些问题都值得进一步研究。

现实中，人工智能侵权责任的归责原则，可能更多涉及危险责任或无过错责任。例如无人驾驶汽车致害，无论从产品责任还是机动车交通事故责任上看，都似乎倾向适用无过错责任，但这远远不是结论。未来需要考虑的是，人工智能技术的运用，其本身是否属于高度危险作业（如无人机），从而决定了是否适用高度危险作业致害责任。

算法歧视指的是数据的选用和人的主观想法会在机器学习中产生偏见。例如，谷歌曾错误地将黑人照片打上“大猩猩”的标签，雅虎旗下的 Flickr 也曾错将黑人的照片标记成“猿猴”。微软公司的 AI 聊天机器人 Tay 上线，但在和网民聊天时却被“教坏”，被灌输了许多脏话甚至是种族歧视思想，变成了一个“不良少女”，因而上线不到一天就被微软公司紧急下线了。要解决算法歧视问题，需要从技术和制度切入，保证其公平性、透明性和可追责性。

当前，人工智能侵权责任中的因果关系、过错等要件的判断也变得日趋复杂。一些APP“大数据杀熟”和“算法歧视”，代码的不透明，以及算法本身的自我学习和适应能力，使得简单地“将算法歧视归责于开发者”的责任认定变得非常困难。

针对人工智能带来的新问题、新挑战，在法律制度的研究方面未雨绸缪，将为以后的司法实践赢得主动。人工智能已经到来，只是在生产生活的各个领域分布不均。我们不应等到未来分布均匀、人工智能已完全融入生产生活的方方面面时，才想起从法律角度对其进行规范。

※ 活动2　维护人工智能社会化应用的伦理道德

按表5.2.6所示的主题划分小组，在老师的指导下，设计出高效的搜索关键词，如自动驾驶的伦理困境；在老师的推荐下，选择较权威的官方网站，开展数字化学习。通过数字化学习，了解人工智能的伦理道德问题。

要求每个小组针对上述主题开展协作学习，并完成相关主题的知识获取、筛选与整理，然后以小组为单位在班级中进行分享与讨论。

上述活动完成后，请填写表5.2.6。

表 5.2.6　人工智能的伦理道德问题活动记录表

序号	问题	讨论	答案
1	汽车高速行驶中，主道上有 5 人，侧道上有 1 人，假如事故不可避免，救 1 人还是救 5 人		
2	假如交通事故不可避免，救乘客还是救路人		
3	是否愿意购买倾向首先牺牲驾驶员的自动驾驶车辆		

霍金的担忧

霍金在其演讲《让人工智能造福人类及其赖以生存的家园》中指出，关于人工智能，一个短期的担忧在于无人驾驶方面，涉及从民用无人机到自主驾驶汽车。例如，在紧急情况下，一辆无人驾驶汽车不得不

在小概率的大事故和大概率的小事故之间进行选择。另一个担忧在致命性智能自主武器方面。它们是否该被禁止？如果是，那么“自主”该如何精确定义？如果不是，任何使用不当和故障的过失应该如何问责？还有另外一些担忧，如由人工智能逐渐可以解读大量监控数据引起的隐私担忧，以及如何管理因人工智能取代工作岗位带来的经济影响。

长期担忧主要是人工智能系统失控的潜在风险。随着不遵循人类意愿行事的超级智能的崛起，那个强大的系统可能会威胁到人类。这样错位的结果是否有可能？如果是，这些情况可能会以怎样的形式出现？显然，这是我们必须重视的问题，我们应该投入足够的关注和研究。

红十字国际委员会将自主武器定义为“可以根据自身所部署的环境中不断变化的情况，随时学习或调整运转的武器。真正的自主武器能够在无人干预或操控的情况下搜索、识别并使用致命武力攻击包括人类在内的目标（敌军战斗员）”。也有人将自主武器称作“杀手机器人”，其“是具有某种形式人工智能的移动系统，能够在无人控制的情况下在动态环境中运行”。但自主武器的研发带给了人们更多的担忧，这些担忧主要集中在攻击目标方面，人们担心自主武器无法区分军事目标和平民，从而造成对人道法的违反。

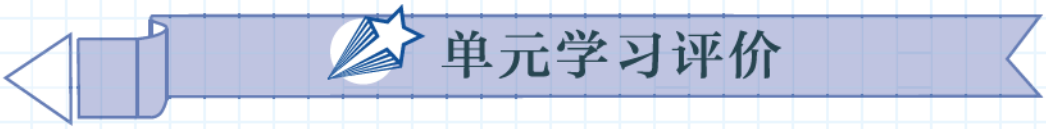
拓展练习

2017年11月，一款名为Stinger的小型机器人在联合国《特定常规武器公约（CCW）》会议上亮相。这款机器人只有手掌心大小，却是一款能携带炸药、精准识别目标对象的人工智能飞行武器。与以往很多不能准确定位目标的武器不一样的是，Stinger小型机器人搭配有广角镜头、战术传感器、面部识别技术和反应比人类快100倍的处理器，可以全自动地通过步态、性别，甚至种族等特征来识别目标，还特地针对反狙击手设置了随机运动模式。

国际范围内应该禁止致命性自主武器系统的使用吗？请分成“正方”与“反方”两组，就此主题进行辩论，填写表5.2.7。

表5.2.7 是否禁止致命性自主武器系统的使用辩论活动记录表

正方主要观点	反方主要观点	结论



单元学习评价

通过本单元的学习，我们体验了声纹锁、指纹锁和变声软件的安装与使用，了解了其中的利弊，了解了人工智能的巨大价值和潜在威胁。你是如何看待智能系统的？通过智能系统的应用体验，你是否了解了社会智能化所面临的伦理及安全挑战？是否知道了信息系统安全的基本方法和措施？是否了解了人工智能对人类社会的巨大价值？是否了解了人工智能的未来发展和潜在威胁？是否知道如何遵守人工智能社会化应用的法律法规？是否知道如何维护人工智能社会化应用的伦理道德？请参加小组交流并反思，开展自评或小组评价。

1. 指纹认证，是用_____作为生物特征来验证用户身份的技术，不容易被仿冒，也_____。

2. 华丽琴鸟公司利用_____便可准确复制用户声音，等到该技术对公众开放之后，_____便不应再被视作验明正身的证据。

3. 人工智能入侵检测系统会根据其机器学习模块在_____的训练结果，对接收网络数据包的信息进行_____，判断是否是入侵行为或异常现象。

4.《人工智能的恶意用途：预测、预防和缓解》给出的5点建议是：

（1）_____；

（2）_____；

（3）_____；

（4）_____；

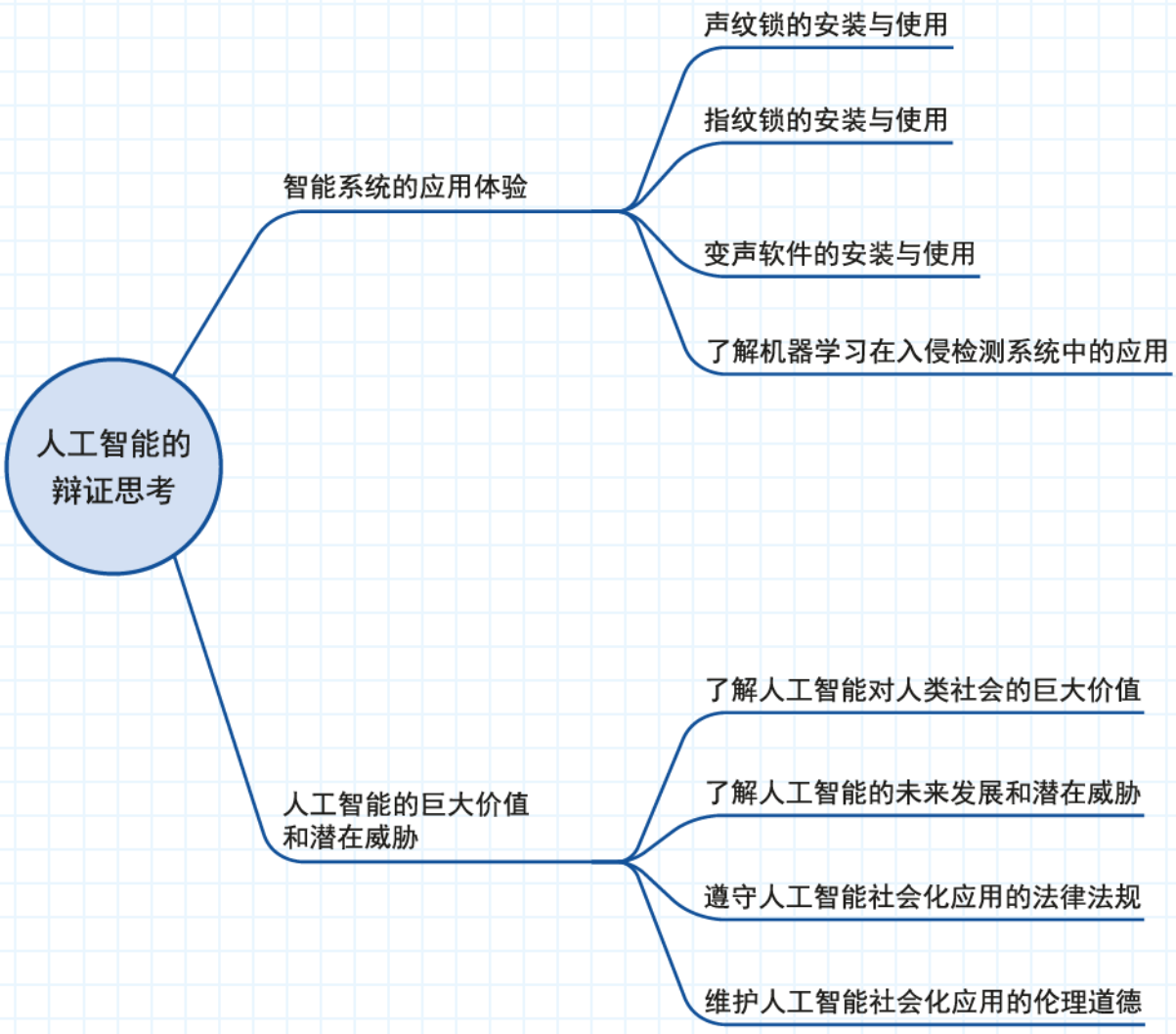
（5）_____。

5. 针对人工智能带来的新问题、新挑战，在法律制度的研究方面应该_____，为以后的司法实践_____。

6. 人工智能的具体行为受程序控制，发生侵权时，应该是由所有者还是研发者承担责任？_____

7. 应该在国际范围内禁止致命性自主武器系统的使用吗？_____

单元学习总结


人工智能的辩证思考
智能系统的应用体验
声纹锁的安装与使用
指纹锁的安装与使用
变声软件的安装与使用
了解机器学习在入侵检测系统中的应用
人工智能的巨大价值和潜在威胁
了解人工智能对人类社会的巨大价值
了解人工智能的未来发展和潜在威胁
遵守人工智能社会化应用的法律法规
维护人工智能社会化应用的伦理道德

后　记

为全面落实立德树人根本任务，着力发展学生的核心素养，根据《普通高中课程方案（2017年版）》的精神，我们按照《普通高中信息技术课程标准（2017年版）》的要求对高中信息技术教科书进行了修订。

本书的修订由张剑平、余燕芳、林斌、陈美锭、白晓东等直接参与，李艺、董玉琦、解月光、李冬梅、张义兵等在整体设计的过程中给予指导，后期陈斌、李千目和柏宏权教授审阅了本书修订稿并提出了宝贵意见。在此，我们对所有关心、支持本书编写与修订的专家、学者和老师们表示衷心的感谢。

本书选用了一些图片和文字资料，对相关作者和出版社，我们一并表示诚挚的谢意。

编者

2019年6月