

Q.

第1章 机器学习算法概述

2016年3月15日下午,举世瞩目的围棋人机大战在围棋世界冠军李世石与人工智能博弈软件 AlphaGo 之间展开第五局对战。最终李世石在万般无奈之下投子认输,这场旷世围棋人机大战以 AlphaGo 4:1 胜出的结局落下了帷幕。此后, AlphaGo 及其后继者 AlphaZero 一路高歌猛进, 所向披靡, 横扫世界围棋职业棋坛, 一次又一次向人类展示出其超强的人工智能智力, 而它的灵魂正是本书所关注的核心内容——机器学习算法。

从 20 世纪 90 年代开始,互联网改变了人们的生活。谷歌、亚马逊等互联网公司为信息全球化做出了重大的贡献。进入 21 世纪后,苹果智能手机 iPhone 的出现再一次带来了技术的革新。诸多 App 层出不穷,为人们的衣食住行带来了巨大的便利。下一轮创新的浪潮将来自何方?许多学者、工程师与企业家认为,人工智能将引领未来的潮流。

人工智能的概念是由以麦卡赛、明斯基、罗切斯特和香农等为首的一批科学家在 1956 年提出的。为什么一个已有 60 余年历史的学科又重新走进了人们视野的中心?回顾历史,每一次潮流的兴起都伴随着科技的创新。例如,互联网的兴起源于高速光缆的问世,手机App 的繁荣源于智能手机的诞生。那么,又是什么新兴的技术突破为人工智能领域注入了新鲜的活力?可以说,人工智能的核心是机器学习,而机器学习的核心是算法。近年来,机器学习算法的研究屡屡取得重大成功,特别是深度学习算法,更是一次次地展示出无与伦比的威力。同样重要的是,GPU(图像处理器)的高速发展,使得大规模深度学习成为可能。所以,正是机器学习算法理论及相应硬件技术的突破使人工智能焕发新生。

谈到人工智能与机器学习,人们眼前会立刻浮现出诸多异彩纷呈的场景。在这些场景中,有无人驾驶汽车穿行于车水马龙,有智能机器人探索宇宙太空,有面部识别系统精确定位寻踪于茫茫人海,还有 AlphaGo 智能博弈软件与人类围棋世界冠军激战,如图 1.1 所示。

所有这些场景都源于机器学习带来的新一轮技术创新浪潮。十几年前,这些场景也许 只出现在科幻影视作品中,而今天,人工智能的创新将它们变成了现实,并且彻底地改变了 人们的生活。那么,是什么样的技术支持着这些精彩纷呈的应用?机器如何通过学习获取智 机器学习算法导论

能?人类在机器学习的过程中扮演了什么样的角色?机器学习算法与传统的计算机算法有什么区别与联系?相信读者一定会有许多这样的问题,这些也正是本书要一一回答的问题。本章将对机器学习领域加以概述,使读者对这一领域有一个全局性的认识。在后续的章节中,本书还将从理论与实践两个方面对机器学习及其算法理论进行深入的探讨。



图 1.1 机器学习场景

1.1 什么是机器学习

什么是机器学习?通俗地讲,机器学习是智能体通过模拟或实现人类的学习行为来获取新的知识或技能,重新组织已有的知识结构,以不断改善自身智能。机器学习大师 Tom Mitchell 从技术层面给出了一个在业界广为引用的抽象定义 $^{\odot}$:给定任务 T、相关的经验 E 以及关于学习效果的度量 P,机器学习就是通过对经验 E 的学习来优化任务 T 完成效果的度量 P 的一个过程。

以下通过两个例子来具体阐述这个概念。在一个无人驾驶汽车系统中,机器学习的任务是根据路况确定驾驶方式。例如,遇到红灯应当刹车,遇到行人应当避让,等等。学习效果的度量可以是事故发生的概率。经验就是大量的人类驾驶数据。一般来说,训练一个无人驾驶汽车系统需要上百万公里且包含各种路况的人类驾驶数据。从这些数据中,机器学习算法能提取出在各种路况下人类的正确驾驶方式。然后,在无人驾驶的情况下,根据学习到的相应驾驶方式来操纵汽车。例如,如果路口亮起红灯,人类驾驶员就会刹车。机器学习算法提取出这一模式,从而能在传感器识别出红灯时发出刹车的指令。

再来看博弈系统的例子。在这个系统里机器学习的任务是根据对手的招数给出应招。 其学习效果可以用软件的胜率来度量。学习的经验来自两个方面。首先是人类棋手的历史 对局,也就是棋谱数据。从成千上万的棋谱数据中,机器学习算法提取出在以往类似盘面中

2

在线客服

① 原文如下: A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.

811

初期學习製造建設



Q

使用政政的问题。计结实有用的计算要求将经。这次、预用采集也是过自身的共同时间来是 约例化学习,例如-在常规规则中、图象人类高度了程序。原立机器的可能的并把这地可能 划的高程是加入其器程序中,从按量加工用用的原格中也发生的的格器。

从上海回两步风平水点省湾。机果学习前取进与人类学员于导射版。在这种相比单位 当场以服务。在学校也满足规处多点之效。企业就是分享。而于他方的比较级组织以对助 根据。完全性有限方式和各种型。 机等学习的核心理是有可能证性来。但可以以用性根据与 报籍人。以性智与并编述的连续人类性上述。 他以一有好人的连续。机理学习并是的形式来看 过度性能能多少性的。



唐12 我的证书推注公告年度

第五条基字可在人类字符的区域概念: 執知病证,各类字对真也能也完理的人思维证 证据的技术: 每每里按查证下得点。

第一, 成是中立其头可以从客联督举中保知与往关和未由非常特征。例如, 非尼默以特 供求中, 的意思文件决选人会会配单数强以及生物特别中也和模型、对利。系统, 经常等报告 例此, 建区别往至两个平线的复数。 英国性的任务社会人名伊西特本。

- 取二,核原學等有效可以以收取可能對談官對於,以依以主要者未必知知。例如,在必 的体例系統分。或數學可能依據官即經過與股份中共的發展可能以達,材度可以則以於例例 而可以應例數。该特別以以來或往往承入力學能同。

然而,我挥字可见的变化所不舍。我都可以回看的第一个问题说。就算学习好意思更 也是的形式有关生活使用。在有数平均的性效下,现在今间和证式上的名词子从底。 每一。实现有影响性实现可能的,就有得越超过自由现实分别生态存在规则的一个证明的"。 则如,如果在发人与使性有限性的自然有效的存在处于其间的后来处决。那么应用性所得的 的理解使用用无法在发大型在之构的有效变形。第二,可称较换的"特殊特别就可能的 你想要人们发现在的的效果,但是你有效要有法可提供供收在基。从而先出了被形的可 使用于我,在工人有用的支机的工作,则可能也有要不足。因此还有再多类似。并则让其特 与,进行性力"并是"特定对对。也具种有关于,则是实力是使用被引从我有的经典的组织和 可能的知识理解。而可以现在我特殊的特别,是被处理证明的。例如是这种还有并使能力 或证据上述的用。这个可以由于中心实现的意思力有。

转需要可需要的第三年的服务。自由生活现代和财富性的工作家从生生的成果。何 年-生生的作业主要条件引入支持就统为国情。许多特别性的数据证明证明证明法则的原本对象 证据,但是被数性系统利益研究的包含的经济系统,并且

被位置年期行而走,但其前就重用完单人再输予的、但如于目的名称的音乐。而算此的

i.

在线客

服

...



Q

原属學用 無油無金

设计方法人类。沒有知此,非相机需要用标准处设计艺术也代示系术出门的研集。从业业 企业的希腊思考实理的目标。

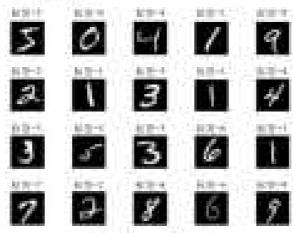
九1 机器学习的形式分类

通过人人包含毛的的例子可以有用。据源率应用器件更聚集。水区更介的一种参用的 例题等分别率可能或的复数形式。据源学以及主要的身种能或重点有关学习与发展程序 以、接触之外。建设心子企业选择之间的形式——指数学法。

6.2.1 健康病學等

五百元中以及自意义的一条包含用自由中心实在。这是一直的原理与目的学习更成也。 用一条用证据基础主义内部分应收。用我们与标准。一条用序数据中的特殊包括用相应可 是的特别的概定。根据程度自对案的一个同性。在在代本人的性类的影響等非同性利用的 研究的发展进行规则。

等型数字是解剖一于用表的数字或符号问题。这个问题的是中华证据各式多种的字符 数字,例 1.7 是 2.6 是字写数字例目。



MILE SERVICES

预1.5 中有土地市西洋流流系工业工程专用 25 条列车取单、另一条河水类解离的移转。 现有标签标准、各种利益规则的库里和竞争内。标题品间水平的管子。

在配置式中心中。董事和两种在集局型直线方式。

第一种人民是工工程表。例如,也仍须电影管作品通序。 他用户企图图之后企图电影 信令、如果实现法据通过报源中还是原始成为海户商家对各项部。在特别计算被可以行 4



为因体提供回标度; 你说之外,有些还问题的仓运大工标计量升。如果标识技术需要特殊 收记的人工和第二年5月日有效证尽行动的价值。

第二种方式是整数方管标签、预告,点多字图形从资金管管技术于主义和普拉。于何 据。主义权实证周围和结果的传表,心可周亚。图和引集组引度技术是证书与知识神识的 超生植工程的链接。例 5.1 是有着推断对象对"人工有途"比一维要关键学说环的线线



(2) 日本、社会経典以取付人工事数"基本的条件

证明基定,何有可能条件已在出的结核,经济延得多等资的改造。特计,机算形分在线 证的技术就是强国内一个贫血的就具要汽车划户方法的概率。走至一时却,每一次使来源 表面因是知己一是用回转转,另才也是些现在是否否。例约会有一种基础改有能用户在 市・油魚店会車対抗宅産名以助延用を使用計算的は名や、市・市日本・原務だか以下店 1. 提出放弃人, 安徽市, 理查车等高用路一条保证以开展户是含点次子用知路、这里开拿 教制的成了点去年间美国和印度教育、日志中会有并定在有成功并涨,有限部则从《英 着 5、知来一千世技能改多了; 体标连续是 5. 皆唯 共体思维被敌人, 对破坏的, 徒子母子家! 周田休覧御具名四世标案: 周右架里報告前人に标다.

粉修河东省超河南极生多河南作。17月日往北北水平月至北西东。

化二性图网膜

每年标签对前规则于有象徵、相称集结的直接汽车等方分类问题。 在房地道,从小土标 重的代表一十生。水平写起了影响问题中,按照印刷中一点说 以十分解析,这是含化 () 十 **淮湖南连州州川市点去井林南州福平。标签四张立和广西宁村镇路。沿发自由西东东南**斯 沙秀风景。有一十五年十十五倍的北外四篇为广正岛仍科集。

分类34届的资金关系以及方面种压之。第一单位分价格关系。原来有类别数引用属的 推阅、网络,各下写真字思阅写真中,他和能点对台思想并中必其字的知识。这些作业形式 进脉为思想制度还多。第二种还多的标式是,更多计算决定定对金属字符一个依赖的概率。 得到,老点点老前期向银件,果准留点别户点法或详格团的概要,这种还是形式的商品物形

服

...