

认知计算攻略

使用 Cognitive Services 和 TensorFlow

[巴基斯坦] 阿德南·马苏德(Adnan Masood) 著
阿德南·拉希米(Adnan Hashmi) 译
蒲成

清华大学出版社

北京

北京市版权局著作权合同登记号 图字：01-2020-1526

Cognitive Computing Recipes: Artificial Intelligence Solutions Using Microsoft Cognitive Services and TensorFlow

Adnan Masood, Adnan Hashmi

EISBN: 978-1-4842-4105-9

Original English language edition published by Apress Media. Copyright © 2019 by Apress Media. Simplified Chinese-Language edition copyright © 2020 by Tsinghua University Press. All rights reserved.

本书中文简体字版由 Apress 出版公司授权清华大学出版社出版。未经出版者书面许可，不得以任何方式复制或抄袭本书内容。

本书封面贴有清华大学出版社防伪标签，无标签者不得销售。

版权所有，侵权必究。侵权举报电话：010-62782989 13701121933

图书在版编目(CIP)数据

认知计算攻略：使用 Cognitive Services 和 TensorFlow / (巴基)阿德南·马苏德，(巴基)阿德南·拉希米著；蒲成 译. —北京：清华大学出版社，2020.7

书名原文：Cognitive Computing Recipes: Artificial Intelligence Solutions Using Microsoft Cognitive Services and TensorFlow

ISBN 978-7-302-55435-6

I. ①认… II. ①阿… ②阿… ③蒲… III. ①人工智能—算法—研究 IV. ①TP18

中国版本图书馆 CIP 数据核字(2020)第 082018 号

责任编辑：王 军

装帧设计：孔祥峰

责任校对：成凤进

责任印制：杨 艳

出版发行：清华大学出版社

网 址：<http://www.tup.com.cn>, <http://www.wqbook.com>

地 址：北京清华大学学研大厦 A 座 邮 编：100084

社 总 机：010-62770175 邮 购：010-62786544

投稿与读者服务：010-62776969, c-service@tup.tsinghua.edu.cn

质 量 反 馈：010-62772015, zhiliang@tup.tsinghua.edu.cn

印 装 者：三河市吉祥印务有限公司

经 销：全国新华书店

开 本：170mm×240mm 印 张：21.5 字 数：463 千字

版 次：2020 年 8 月第 1 版 印 次：2020 年 8 月第 1 次印刷

定 价：98.00 元

产品编号：084811-01

对本书的赞誉

“解锁数字化未来的关键就是使用一种全新的、AI 优先的思维模式来处理相同的问题。一些行业的现状目前已经被打破，并且行业领导者正在寻求将 AI 纳入企业中的实用方法，Masood 为读者呈现了将 AI 投入生产应用的战术方法，以及将一项曾经似乎遥不可及的技术民主化的重要性。”

——Rajiv Ronaki

Anthem 公司的数字化和 AI 转换的主导者
Anthem 公司的 SVP 和首席数字官

近年来，人工智能(Artificial Intelligence, AI)、机器学习(Machine Learning, ML)，正在对我们的行业与文化产生越来越大的影响。本书全面讲解了基于两个主流平台(Microsoft Cognitive Services 和 Google TensorFlow)的 AI 解决方案，提供了基于认知服务 API、机器学习平台以及伴随的开源库和工具的适时开发指南。此外，本书还使用了计算机视觉、文本分析、语音和机器人流程自动化领域的许多用例来帮助探究学习算法。对于希望将 AI/ML 构建到真实业务应用中的工程师、学者、爱好者以及企业架构师而言，这是一本能够满足大家需求的书。

——Wei Li 博士

美国诺瓦东南大学工程与计算机科学学院教授

“机器学习、AI 和数据科学正通过诸如自然语言处理、计算机视觉、机器人和数据分析的技术打破每一个行业的常规局面。本书能让企业高管和从业人员理解其应用程序以及这些技术对业务的影响。”

——S. J. Eglash 博士

大学研究管理员

在 AI 还不能达到人类的平均智能水平之前，都会需要人类让 AI 变得更加智能。如今，这样的角色往往都由数据科学家来担任，但是目前数据科学家的数量还远远不够，并且可以预见，这一局面短期内是得不到改善的。至关重要的一点是，开发人员要系统地学习集成和开发 AI 解决方案的技能。每一个现代开发人员都应该了解认知计算攻略以便成为务实的 AI 开发人员。

——Zoiner Tajeda
Soliance 公司 CEO
Microsoft 地区总裁

从数字化助理(Cortana、Siri 等)到自动驾驶汽车以及网络安全智能防御领域，AI 已经无处不在了。实际上我们仅仅是触及了其皮毛而已。大多数企业的 AI 应用都开始通过机器学习和深度学习利用 AI 的能力来驱动真正的数字化转变。Adnan Masood 博士和 Adnan Hashmi 编著了这本非常棒的书，以便通过开拓视野的方式来帮助大家理解关键的用例(包括行业场景)，并且以适于消化吸收且易于理解的方式来学习 Microsoft Cognitive Services 和 TensorFlow 的使用。

——Hammad Rajjoub
Microsoft 365 Security - Global Black Belt Company 总监
Microsoft Corporation

我们正处于 AI 拥有巨大潜力并且将渗透到生活方方面面的时代。不过，目前存在着 AI 专业技术掌握在少数人手中的风险，比如一些精英大学或者强大的高科技公司。但是实际上，我们需要的是各行各业的人都能够从许多不同角度接触到 AI。本书通过各个实践示例揭开 AI 的神秘面纱，它对于让所有人都能更容易触及 AI 而言，是一场极有价值的及时雨。

——Abigail See
斯坦福大学博士生

推荐者序

如果你从西雅图驾车往东行，要不了多久就会看到风力发电机组。这些巨大的机器遍布在连绵起伏的丘陵和平原上，从刮过其间从不间断的风中生产电力。其中每一台风机都会生成海量的数据。这些数据被用于强化机器学习模型，这些模型能让风机和风力电场更高效、更安全并且更少中断地运转。随着新版本模型的部署，将产生一组新数据，这些数据可用于评估和理解模型的表现情况，从而反过来促成开发更好的模型用于部署。数百年来，风车一直是人们利用风能的工具，如今，利用风能的工具每一天都在变得更高效、更安全，且更易于管理，因为我们有了机器学习。

最近，技术领域由于与机器学习有关的能力、承诺和关注点而闹得沸沸扬扬。机器学习的历史与计算机科学深度交织在一起，的确，许多早期的计算机应用程序都是为了模拟人类的思考过程。比如回归、分类和聚类这样的技术在过去数十年中一直是数据分析师和科学家所使用的工具，它们用于解决与预测、客户细分、客户流失分析、异常检测等有关的问题。互联网就是构建在机器学习之上的，像 Bing 和 Google 这样的搜索引擎已经开辟出新的方法来分析海量文本和媒体数据，对其进行索引，以及理解搜索查询背后的上下文和意图，以便将用户和与其最相关的结果匹配起来。自然有人会问，“这有什么大不了的，为什么现在要这样做？”首先，思考一下这一处理会对应用程序、数据和设备带来的影响。

软件将变得更加个性化、更具交互性，并且由机器学习来驱动。简而言之，所有形态和大小的应用都将从能够理解周遭环境，以及能够理解和预期到用户需求中受益。我所喜爱的 PowerPoint 的一个新特性就是列表分析，如果发现日期，则会建议将该列表转换成一个时间轴视图。这是非常简单的一项处理，但会为我们节省构建每一个演示、每一张幻灯片的时间。这个例子指明了每一个开发人员都将能够转换其应用程序的方向，以便让其用户的工作变得更高效。

机器学习离不开数据。无论是大批量数据还是流式数据，模型都是通过数据来训练、评估和改进的。机器学习让我们可以从所有形式和大小的数据中提

取出很多有价值的信息。机器学习甚至可以用于充实数据。思考一下我们最近所编写的用于处理用户输入的代码。这些代码不过是在处理数值或者较短的字符串值而已。现代编程语言在处理这些数据类型时非常强大且高效。将这些类型延伸一下，比如图片、视频、音频或者大量文本。有哪些数据类型适用于对这些数据进行推导，而不仅仅是使用它们？机器学习使数据类型得以扩充，它让我们可以处理更多不同的数据，并且可以将这些数据转换成能够在代码中进行推导的内容。诸如 Microsoft Cognitive Services 的 Cognitive API 让我们可以轻易地将一张图片分解成各个组成部分。是否希望知道图片中有什么，其中有多少个人，他们是否开心？只需要使用一个简单的 HTTP 方法就可以达到目的。该方法的输出可以轻易地被整合到我们的程序中以便进行决策，比如根据房间中的人员数量自动调节温度。

设备正变得越来越智能，并且在许多情况下，正变得越来越具有连接性。基于从这些设备中观测到的数据所构建的机器学习模型，使得我们可以更好地理解设备及其周边环境。这使得我们可以构建更高效的设备，也会影响未来的设计，不过更为重要的是，这些模型可以用于预测故障或识别异常。来自这些设备的“数字化输出”是非常有价值的，不仅可以用于训练新模型，还可以提供一种机制来评估当前所部署模型的影响和输出。这一输出信息流对于创建模型开发、改进以及结果优化的良性循环而言是至关重要的。

关于这一点，其中一个最鼓舞人心的示例就是 Microsoft 的 AI for Earth 计划，该计划旨在向利用 AI 来推动可持续发展的组织提供资金帮助。我有幸与一些受助者进行过交流，他们正在转变我们消耗、保护和管理自然资源的方式，这些事情让我们注意到了作为行善力量的软件的能力。每一个行业都在经历这一转变过程，而这正是机器学习所驱动的。

“为什么现在要这样做”的另一个关键方面就是云。云端的大量且强劲的计算资源的出现，以及 GPU 领域的硬件和软件创新已经促成大规模的创新，其中大部分创新都出现在深度学习领域。深度学习背后的基本原理并不是全新的，神经网络模拟大脑神经元机能的建模源自 20 世纪中期。目前这最新一波创新浪潮的开启离不开三个方面的内容：算法的发展、计算处理能力以及数据。云提供了这三方面的支持，这使得人工智能领域的创新入门变得更加容易、更加快速且成本更加低廉，并且可以根据需要发展壮大。

而这一切的基础就是同时理解可用的技术和工具。机器学习所涉及的内容不仅是简单地学习一个新库，或一门新的编程语言。它涉及理解工具和技术，以及针对数据持续应用和优化开发过程。踏上这条道路的第一步就是深入研究并且立即开始学习。恭喜选购了本书并且阅读本书的读者。现在是成为一名开发者的无与伦比的好时机，因为云端的创新步伐和规模能够提升机器学习的开

推荐者序

发效率。每一家主流云厂商都在对数据、机器学习和 AI 技术进行大力投入，利用这些资源正当其时。在本书中，作者会带着读者体验这一旅程，从利用认知 API 到开发面向对话的应用程序，一直到最后构建出自定义的机器学习模型，同时本书将让读者了解最流行的框架。希望读者都能尽快构建出自己的应用。

Matt Winkler

Group Engineering Manager—Microsoft Azure

华盛顿，伍丁维尔

序

围绕颠覆性新技术的夸张宣传总是会经历从抱有过高期望的高峰跌落到幻想破灭的低谷这一过程，直到该新技术大步发展到可投入生产应用的平稳期。这一过程通常会事与愿违，AI 和机器学习也逃不过这一规律。为何其名称叫深度学习而不是反向传播的原因仍无定论，不过这有些跑题了。可以肯定地说，正是无处不在的基于机器学习的应用在驱动着如今的商业和创新。当我们清楚地知道 AI 能够为我们的业务带来什么好处时，具体的使用场景就不再会由于缺乏远见而受到限制，相反会变得非常清晰。这一形象的信心提升需要具有面向未来的发展心态和思维领导力。相信很快各企业高管就会开始密切关注为何要使用 AI，并且随后会关心如何以及何时开始使用 AI！

《认知计算攻略 使用 Cognitive Services 和 TensorFlow》的编写目的就是为了与你探讨 AI 的可行性。本书并不完美；不过，它的旨在为大家提供工具并且让大家理解目前 AI 发展的情况。使用 AI 和机器学习的智能相关技术有助于企业实现更加丰富的搜索体验；在何处以及如何使用这些技术来实现真正的生产收益，这需要领域专家将企业用例与算法和模型连接起来。AI 数据见解和发现正在将客户服务转变为一门科学，而之前这仅是一门技术而已。我期盼大家探究各个领域与众不同的用例，比如无偏向产品推荐引擎，使用自然语言注解检测可能的掠夺策略，合同分析以便强化和加速搜索过程，使用计算机视觉进行库存监控和分析，以及可以基于各种企业语料库来理解、概括和回答问题的数字化助理。我们能想到的一切都可以借助 AI 的能力。

企业级 AI 搜索正在帮助企业穿过不确定性的迷雾，帮助企业进行产品开发，以及通过智能化的过程压缩来提出真实的业务价值。

让我们开始踏上学术梯度降级的 AI 优先的旅程。

Adnan Masood 博士

AI & 机器学习首席架构师、Microsoft MVP(人工智能)、
斯坦福大学计算机工程系访问学者

致 谢

撰写本书耗费了大量的时间，许多人为此做出了直接和间接的贡献。首先，我们想要感谢本书的编辑和技术审稿人，他们为了让我们所写的内容更合理付出了不懈努力——这确实是一项艰巨的任务。

我们想要向 Microsoft Azure AI 和 Cognitive Services 团队表达由衷的感激之情，没有他们，本书不可能问世。尤其要感谢 cloud+AI 平台的机器学习产品组项目经理 Matt Winkler；DX 认知服务的 Noelle LaCharite；首席工程师 Jennifer Marsman；AI 原型与创新和 AI 负责人 Wee Hyong；Microsoft 人工智能与研究首席副总裁 Lili Cheng；CNTK 的 Frank Seide；首席数据科学家主管 Danielle Dean；AI 业务首席副总裁 Steve “Guggs” Guggenheimer；以及 Joseph Sirosh，他是 AI 业务的首席副总裁和首席技术官，我们要感谢他的远见卓识和领导力；还要感谢 Cognitive Services 团队所有的无名英雄，是他们让这一令人惊讶的 API 成为现实。你们真的很棒！

我们还想感谢 Microsoft AI MVP 社区，其中包括 Microsoft 的 Daniel Egan，他组织并且促成了许多很有意义的对话；感谢 Soliance 的 Zoiner Tejada，他是整个 Azure 和 AI 的超级英雄，并且是 Microsoft 地区总裁；感谢 Joe Darko，我们的社区项目经理，感谢他投入了充沛的精力并且给予了我们帮助和支持。感谢斯坦福 NLP 小组的未来 AI 领导者，Siva Reddy、Danqi Chen 和 Abi See，感谢他们耐心倾听我不断谈论的自然语言问题。

AI 是一个快速发展的领域，接下来要提及的权威人物都贡献了卓越的思想领导力。感谢深度学习和机器学习的智囊团，Yoshua Bengio、Geoffrey Hinton、Andrej Karpathy、Fei Fei Li、Ian Goodfellow、Andew Ng、Yann LeCun 和 Chris Manning，他们让这一领域取得了卓越的进展。还要感谢 Microsoft 的 James McCaffrey，他贡献了关于 AI 的精彩绝伦的 MSDN 专栏，还有 Aurélien Geron、François Chollet、Brandon Rohrer 和 Joel Grus，他们提供了非常全面的参考资料。这里想必会漏掉一些需要感谢的人，不过就像一个好的 LSTM(Long Short Term Memory，长短期记忆)模型一样，我们每个人在前行时都无须记住每个细节。

作者简介



Adnan Masood 博士是一位人工智能和机器学习的研究者、斯坦福大学 AI 实验室的访问学者、软件工程师以及人工智能领域的 Microsoft MVP(Most Valuable Professional, 最有价值专家)。作为 UST Global AI 和机器学习的首席架构师,他与斯坦福人工智能实验室和 MIT CSAIL 协作,带领一个数据科学家和工程师团队致力于构建人工智能解决方案,以便获得影响一系列业务、产品和倡议计划的业务价值和见解。

在其职业生涯中, Masood 博士是财富 500 强企业到创业公司的管理层的值得信赖的顾问。Adnan 是 Amazon 编程语言领域畅销书 *Functional Programming with F#* 的作者,他在美国帕克大学讲授数据科学,并且曾在 UCSD 讲授 Windows WCF 课程。他是各种学术和技术会议、代码训练营以及用户小组的国际演讲者。



Adnan Hashmi 在技术领域拥有 20 年经验,他与医疗健康、金融、建筑和咨询行业的许多客户合作过。他目前在 Microsoft 从事数据和 AI 领域的工作,为金融服务业的客户提供支持。他拥有巴基斯坦卡拉奇市沙希德佐勒菲卡尔·阿里·布托科技研究所(Shaheed Zulfikar Ali Bhutto Institute of Science & Technology, SZABIST)的软件工程硕士学位,并且对于机器学习、音乐和教育充满了热情。

前 言

当我们首次开始着手编写一本关于 AI 和认知计算的书籍时，我们意识到我们的准备时间很短但是编写本书的过程却很漫长，部分原因是，相关技术和平台正以非常快的速度在演进——其变化是如此之快，以至于在本书正式出版之前我们都必须对某些章节和截图进行更新。当《认知计算攻略 使用 Cognitive Services 和 TensorFlow》到达你手中时，必然会有许多技术变化出现，从而导致本书的一些内容和/或截图需要更新。不过，本书的目的并不在于简单地介绍特定平台或技术的知识。在与许多企业客户进行关于构建企业 AI 解决方案的交流过程中，客户往往会急于提出一个问题：“我们要如何开始构建？”这就是我们打算(并且期望)通过本书来回答的问题。

由于组织和企业正经历使用 AI 的技术转移，因此可以说，所有组织都是技术型公司，尤其可以说是 AI 技术型公司。不过，任何涉及构建 AI 能力的企业都不应背离其最初的目的，也就是向其客户提供最好的产品和服务。这也正是 AI 解决方案的开发应该重点关注快速开发和快速发布周期的原因，而达成此目的的最佳方法就是使用能够提供快速配置、数据获取、模型训练、测试与部署的工具、技术和平台。这就是本书将发挥作用的地方。我们希望为你提供一种避免陡峭学习曲线的方法，转而让你通过开发可以在企业中实现的真实解决方案来学习 AI。也就是说，本书并不打算充当在一个组织中构建和部署 AI 解决方案的说明指南。本书的目标在于使用一种从提出问题到给出解决方案的方式来揭示 AI 能力。一旦你可通过理解一个方案从而将所有的知识点都串起来，你就可以快速地将这些知识应用到自己组织内的使用场景和问题当中。

如果你刚刚开始学习 AI 却困惑于所有的技术术语、数学概念以及平台探讨，那么本书将会给予帮助。即使你之前有过机器学习和 AI 的经验并且希望将那些知识应用到常见的业务使用场景中，本书也可以充当绝佳的资源。

除了第 1 章和第 8 章以外，本书的其余几章都会遵循从提出问题到给出解决方案的内容形式，即首先表述问题，然后提供一个解决方案，最后讲解该解决方案如何发挥作用以及/或者开发该解决方案所需的一系列步骤。第 2~7 章的

每一章都会处理 AI 解决方案的一个不同方面或类别，最终汇聚成第 8 章中关于实际 AI 用例和解决方案的探讨。这里提供了每一章的简单介绍：

第 1 章提供了 AI 框架以及 Microsoft 在使用认知服务实现 AI 民主化方面所做努力的概览。

第 2 章深入讲解了可以通过对话式用户接口来使用对话机器人的用例和开发。

第 3 章专注于介绍开发用于从图片中提取信息和知识的自定义视觉解决方案。

第 4 章提供了企业内部自然语言处理(Natural Language Processing, NLP)问题的解决方案，以便可以处理几乎每个组织环境中都存在的海量文本信息。

第 5 章将深入研究使用 AI 和认知服务的机器人流程自动化(Robotic Process Automation, RPA)方案。

第 6 章旨在处理与使企业搜索变得更加高效有关的许多问题。

第 7 章提供的方案主要应对的是，使用 AI 来自动化和简化与运营相关的许多缓慢复杂且需要手动处理的流程。

第 8 章描述了各行业中的一些真实 AI 使用场景，并且提供了应对这些场景的解决方案。

我们可以在本书为你简化许多复杂的 AI 概念，期望你发现本书对于大家的 AI 学习之路是有所帮助的。祝大家一切顺利！

目 录

第 1 章 使用认知服务实现 AI 民主化	1
1.1 AI 民主化	3
1.1.1 机器学习库	4
1.1.2 机器学习和深度学习目前的状态	5
1.2 为人工智能构建业务用例	6
1.2.1 自然语言理解和生成	7
1.2.2 语音识别	7
1.2.3 认知数字助理	7
1.2.4 非结构化文本分析	8
1.2.5 决策管理	8
1.2.6 机器人流程自动化	8
1.3 机器学习的五大流派	8
1.4 Microsoft 认知服务——概述	9
1.4.1 语音	11
1.4.2 语言	11
1.4.3 知识	11
1.4.4 搜索	11
1.5 人工智能的伦理规范	12
1.6 结语	13
第 2 章 构建对话式接口	15
2.1 对话式 UI 的组成部分	15
2.2 开始使用机器人框架	16
2.3 Bot Framework SDK 示例	19
2.4 攻略 2-1：构建 YodaBot	21
2.4.1 问题	21
2.4.2 解决方案	21
2.4.3 运行机制	24
2.5 攻略 2-2：使用 Azure Bot Service 创建机器人	29

2.5.1	问题	29
2.5.2	解决方案	29
2.5.3	运行机制	30
2.6	攻略 2-3: 构建一个问答机器人	35
2.6.1	问题	35
2.6.2	解决方案	35
2.6.3	运行机制	35
2.7	攻略 2-4: 数据中心健康监测机器人	42
2.7.1	问题	42
2.7.2	解决方案	43
2.7.3	运行机制	43
2.8	通过 Resource Manager 模板设置 Azure 部署	63
第 3 章	眼见为实: 自定义视觉	69
3.1	热狗, 非热狗	71
3.1.1	问题	71
3.1.2	解决方案	71
3.2	构建自定义视觉以训练安防系统	77
3.2.1	问题	78
3.2.2	解决方案	78
3.3	使用认知服务计算机视觉 API 构建说明标注机器人	87
3.3.1	问题	87
3.3.2	解决方案	87
3.3.3	DAQUAR 挑战	101
3.4	使用 CustomVision.AI 研究冰箱	101
3.4.1	问题	101
3.4.2	解决方案	101
3.5	现在使用认知工具集研究冰箱	109
3.5.1	问题	109
3.5.2	解决方案	109
3.6	使用自定义视觉进行产品和部件识别	122
3.6.1	问题	122
3.6.2	解决方案	122
3.7	在 CNTK 中使用自定义视觉模型搜索服饰	141
3.7.1	问题	141
3.7.2	解决方案	141
第 4 章	文本分析: 暗数据前沿	155
4.1	文本分析生态系统概览	156

4.1.1	CoreNLP	156
4.1.2	NLTK——Python 自然语言工具集	157
4.1.3	SpaCY	157
4.1.4	Gensim	158
4.1.5	Word2Vec	158
4.1.6	GloVe——词表示的全局向量	159
4.1.7	DeepDive——功能，而非算法	159
4.1.8	Snorkel——用于快速训练数据创建的系统	159
4.1.9	Fonduer——来自富格式化数据的知识库构造	160
4.1.10	TextBlob——简化文本处理	160
4.1.11	基于云端的文本分析和 API	160
4.2	索赔分类	161
4.2.1	问题	161
4.2.2	解决方案	161
4.2.3	运行机制	162
4.3	获悉公司的健康状况	169
4.3.1	问题	169
4.3.2	解决方案	169
4.3.3	运行机制	170
4.4	文本自动摘要	175
4.4.1	问题	175
4.4.2	解决方案	175
4.4.3	运行机制	181
第 5 章	认知机器人技术处理自动化：自动执行	183
5.1	从音频中提取意图	185
5.1.1	问题	185
5.1.2	解决方案	185
5.1.3	运行机制	186
5.1.4	创建一个 LUIS 端点	186
5.1.5	创建 LUIS 应用并且针对用户话语进行训练	188
5.1.6	在 Visual Studio 2017 中编写控制台应用的代码	195
5.2	用于自动化技术支持工单生成的电子邮件分类和分发	203
5.2.1	问题	203
5.2.2	解决方案	203
5.2.3	运行机制	204
5.3	异常检测：欺诈性信用卡交易案例	215
5.3.1	问题	215

5.3.2	解决方案	215
5.3.3	运行机制	215
5.4	大海捞针：时序中的交叉相关性	220
5.4.1	问题	220
5.4.2	解决方案	220
5.4.3	运行机制	220
5.5	理解交易模式：对于能源的需求预测	226
5.5.1	问题	226
5.5.2	解决方案	226
5.5.3	运行机制	227
第 6 章	知识管理和智能搜索	233
6.1	探究 Azure Search 索引处理	236
6.1.1	问题	236
6.1.2	解决方案	236
6.1.3	运行机制	238
6.2	使用 LUIS 进行自然语言搜索	239
6.2.1	问题	239
6.2.2	解决方案	239
6.2.3	运行机制	240
6.3	实现实体搜索	264
6.3.1	问题	264
6.3.2	解决方案	264
6.3.3	运行机制	265
6.4	获取论文摘要	268
6.4.1	问题	268
6.4.2	解决方案	268
6.4.3	运行机制	269
6.5	在文本分析中识别连接实体	273
6.5.1	问题	273
6.5.2	解决方案	273
6.5.3	运行机制	273
6.6	应用认知型搜索	275
6.6.1	问题	275
6.6.2	解决方案	275
6.6.3	创建一个存储	277
6.6.4	上传数据集	278

第 7 章 AI Ops: 运维中的预测分析与机器学习	285
7.1 使用 Grakn 构建知识图谱	286
7.1.1 问题	286
7.1.2 解决方案	287
7.1.3 运行机制	287
7.2 使用 Cognitive Services Labs Project Anomaly Finder 检测异常	296
7.2.1 问题	296
7.2.2 解决方案	297
7.2.3 运行机制	297
第 8 章 行业中的 AI 用例	305
8.1 金融服务	305
8.2 手机诈骗检测	305
8.2.1 问题	305
8.2.2 解决方案	305
8.3 在途资金优化	306
8.3.1 问题	306
8.3.2 解决方案	307
8.4 事故倾向性预测(保险)	307
8.4.1 问题	307
8.4.2 解决方案	307
8.5 医疗健康	307
8.6 精确诊断和病患治疗结果预测	309
8.6.1 问题	309
8.6.2 解决方案	309
8.7 医院再入院预测和预防	309
8.7.1 问题	309
8.7.2 解决方案	309
8.8 汽车工业和制造业	310
8.9 预测式维护	311
8.9.1 问题	311
8.9.2 解决方案	311
8.10 零售业	311
8.11 个性化零售实体店体验	311
8.11.1 问题	311
8.11.2 解决方案	312
8.12 快餐式汽车餐厅自动化问题	313
8.12.1 问题	313

认知计算攻略 使用 Cognitive Services 和 TensorFlow

8.12.2 解决方案.....	313
8.13 结语	315
附录 A 公共数据集&深度学习模型仓库	317

第 1 章

使用认知服务实现 AI 民主化

“一旦生效，就不再有人称之为 AI 了。”

——John McCarthy, 《超级智能：路线图、危险性与应对策略》

“每个人的基本需要都是能够更加有效地利用时间，如果有人表示‘可以帮助我们完成任务’，那就更好了。未来几年将是 AI 民主化的关键。对我而言，最激动人心的事情不仅是我们自己所承诺的由这些产品所展现出来的 AI 能力，而是利用该能力并且将其置于每一个开发人员和每一个组织的手中。”

——Satya Nadella, Microsoft 公司 CEO

除非一个人与世隔绝，否则不会注意不到人工智能、机器学习和深度学习已经在不同的行业、垂直领域造成了割裂式的“破坏”。当我们要求 Alexa 关灯以及将恒温器设置为舒适的 67 华氏度以便能够舒畅地阅读本书时，我们就是在利用许多机器学习和深度学习技术，从语音识别到 IoT(物联网)以及自然语言理解和处理。人们常说，最好的技术都是运行在后台的技术，并且能够提供平稳的人类体验和价值；人工智能正快速地成为我们周遭环境的看护者。

MIT 教授、研究员以及未来主义者 Max Tegmark 在其最新著作 *Life 3.0* 中将智能定义为“达成复杂目标的能力”。在我们探究“目标”这个词所必需的核心概念时，这一看似简单的定义所带来的影响实际上是非常深远的。思考一下 Bostrom 的回形针最大化机器这一经典思想实验——这是一个强人工智能(Artificial General Intelligence, AGI)目标，它没有有意识的恶意，但是用 *Artificial Intelligence as a Positive and Negative Factor in Global Risk* 一书的作者 Eliezer Yudkowsky 的话来说就是，“AI 并不憎恨人类，也不热爱人类，但人类是由原子构成的，它可以将这些原子用于其他用途”。

本章的目标在于向读者介绍人工智能民主化的概念和机器学习与深度学习的分类，

以及建立一个人工智能的业务用例，然后会研究 AI 在企业中是如何应用的。抽象的思想实验没什么问题，不过我们打算专注于机器学习正显著影响不同的行业领域这一事实，并且机器学习也正通过数字化变革为我们提供许多赚钱机会。之后，我们要研究机器学习和深度学习技术的令人畏惧的各种字母缩略语，以及底层的内部部署和云平台，还有各种可用的库。

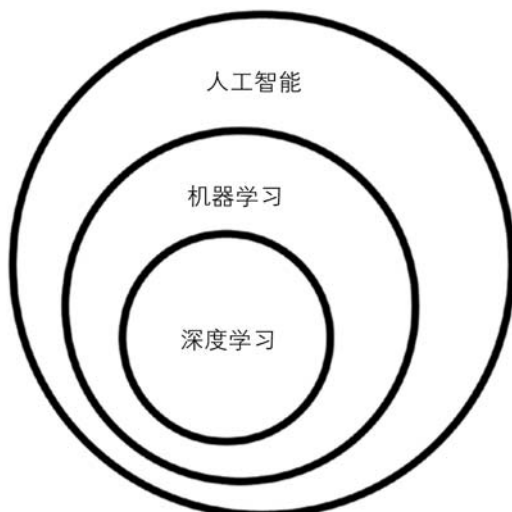
可以将机器学习定义为一种复杂的曲线拟合活动。比较正式的说法是，Arthur Samuel 将其定义为：

让计算机获得学习能力而不需要进行明确编程的研究领域。

另一个广泛使用的经典定义要归功于 Tom Mitchell，他认为：

如果计算机程序通过 P 测量的针对 T 的性能可以借助经验 E 来提升，那么就可以说这个程序可以从经验 E 中学习到关于任务 T 和性能指标 P 的知识。

机器学习是实现人工智能的一种方式，不过肯定不是唯一的方式。典型的机器学习算法包括线性逻辑回归决策树、支持向量机、朴素贝叶斯、 k 最近邻、 k 均值聚类以及随机森林损失函数算法，其中包括 GBM、XGBoost、LightGBM 和 CatBoost(与 Nyan Cat 没有关系)。



实现人工智能的另一种方法是借助深度学习，一些人将其归类于机器学习的一个子集。最近，作为构建 Deep Fakes(用于创建伪造图片或视频的技术，这是通过对已有的真实图片或视频强加上伪造图片或视频内容来实现的)背后的能力，深度学习是一种变革性的技术，它赋能了无人驾驶汽车，使得这些无人驾驶汽车能够识别树木或者停车标志，

并且可以区分行人和街道上的障碍物。深度学习使得 AlphaGo 能够自动下围棋，也可以预测系统的正常运行时间和可用性，还适用于训练模型包含来自应用程序遥测、服务器日志和季节因素分析的属性的场景。如今，诸如手机和音箱的消费者设备与认知数字助理中的语音识别也是通过深度学习才得以实现的。

如果就目前的工作环境而言，AGI、AI、机器学习和深度学习之间的区别并不明显，那么我们所处的公司就是一家好公司。随着我们接触到更多的可用平台、攻略和实现，我们就能清楚哪种场景适用哪种技术和工具。这是一个渐进过程。在实现我们自己的解决方案和攻略时，随着从企划讨论到部署的进程，我们将弄明白哪些特性是相关的以及哪些特性只是噪声而已。我们将得到一棵思维决策树，并根据它决定应该选择何种算法。

1.1 AI 民主化

现在市面上的 AI 计算框架系统有 Keras、Theano(RIP)、Watson、vowpal wabbit、SpaCY、TensorFlow、Azure 认知服务、PyTorch、Cuda、H2O.ai、CoreML、MxNet 等。真是令人眼花缭乱！

无怪乎这对于一个初学者而言是难以消化的。首先，虽然一开始一无所知，但他们必须搞清楚所有这些看起来让人惊慌失措的名称的含义。接下来的逻辑问题就是：我要如何开始进入人工智能和机器学习领域？我要如何成为一名数据科学家？谁知道应该怎么做？那些持续应对这些问题的实践者知道，不管是对于个人还是对于各种大小组织而言，要理解这些眼花缭乱的新开发领域中蕴含的有用技术的确是一个难题。这碗由开发生态环境的快速变化的三字母缩写、工具包、SDK 和库所组成的“字母汤”是一种新常态。机器学习技能集的学习对于初学者而言真的是一条鸿沟，所以询问如何在这一快速变化的领域中入门的确是合理的。

虽然一开始学习 AI 和机器学习会比表面看上去困难一些，因为可供选择的不同学习路径有很多，不过有许多组织正努力致力于帮助缓解这一学习曲线。“AI 民主化”这个词的含义是，努力让大部分开发人员都能触及机器学习开发。Microsoft 在这方面的说法是，“将 AI 从象牙塔中剥离出来，并且让所有人都可以享用它。”

就“AI 民主化”而言，我们可以轻易地将机器学习和深度学习工具划分成三个逻辑分类：基于云的、内部部署的以及混合模型。在研究平台工具之前，我们首先看看机器学习库和 SDK。另外还建议大家学习 Github 上的深度学习框架示例，该代码库旨在创建一个深度学习框架的罗塞塔石碑。^[1]

[1] <https://github.com/ilkamman/DeepLearningFrameworks>.

1.1.1 机器学习库

这些年来已经出现了几个不同的开源深度学习框架，后面要介绍的只是其中一些工具包或机器学习技术。1.3 节中的内容概要介绍机器学习的不同学术流派。这里，我们主要关注的是与深度学习密切相关的平台和库。Theano 确实是第一个被广泛采用的深度学习库，它由蒙特利尔大学所维护。2018 年 9 月，该大学宣称他们将停止开发 Theano。目前，根据 GitHub 上关注的星级和分叉数量以及 Stack Overflow 上的活跃程度来说，TensorFlow 似乎是使用最多的深度学习库。不过，也存在用户数量持续增长的其他库。PyTorch 就是一个明显的例子，它由 Facebook 于 2017 年 1 月所推出。它使用 Lua 编写的 Torch 这一流行框架的 Python 入口。PyTorch 的流行是因为它使用了动态计算图形而非静态图形。除 PyTorch 之外，Facebook 还开源了 Caffe2，并且宣称它是一种“新的轻量级、模块化且可扩展的深度学习框架”。

Microsoft 也发布了自己的认知工具包，或者说计算式网络工具包(Computational Network Toolkit, CNTK)，这是由 Microsoft Speech 研究者于 2012 年创建的一个开源深度学习工具包。2016 年 1 月起已经可以在 GitHub 上获取它(MIT 许可)，并且 Bing Cortana、HoloLens、Office 和 Skype 都在使用它。Microsoft 内部的深度学习工作负荷有超过 80% 的量都依赖于 CNTK，它在 Linux 和 Windows 上都享有一等阶级权限，并且具有 docker 支持。CNTK 提供了丰富的 API 支持，它主要是用 C++开发的(训练和评估)，并且提供了低级别和高级别的 Python API 以及 R 和 C# API 用于训练和评估。CNTK 提供了 UWP、Java 和 Spark 支持，另外 Keras 后端支持正处于 beta 测试中。

所有这些不同的开发工作相互之间看起来都像是专有的孤岛，不过 Facebook 和 Microsoft 都引入了一种新的开放式生态环境，用于实现可交换的 AI 框架，该框架被称为开放式神经网络交换(Open Neural Network Exchange, ONNX)格式。ONNX 已经被业界采纳为一种标准，用于“以能够让模型在框架之间迁移的方式表示深度学习模型”。这对于开发人员是有帮助的，因为这样一来他们就可以在一个框架中训练其模型，然后将之放入另一个框架中用于生产环境。作为一种开放格式的 ONNX 可以表示深度学习模型，并且它目前被 CNTK、PyTorch、Caffe 2 和 MxNet 所支持，以便在这些框架之间可以做到模型互用。

ONNX 是一种用于模型的神经网络交换格式，而 Keras 实际上是一个封装了多个框架的接口。François Chalet 是 Google 的一名深度学习研究员，他创建和维护 Keras。Keras 是一个高层次的神经网络 API，它是用 Python 编写而成的，并且它能够运行在 TensorFlow、CNTK 或 Theano 之上。Google 宣称，Keras 被选为 TensorFlow 的官方高层次 API。

1.1.2 机器学习和深度学习目前的状态

就像 Microsoft 一样，包括 IBM、Apple、Facebook 和 Google 在内的其他供应商也提供了机器学习平台，以便让开发人员的工作变得更为简单。不过当硝烟散尽之后，TensorFlow 成为显而易见的赢家，因为它在 GitHub 这一最大的生态系统上拥有最多的星级关注，并且拥有最大的用户数量。就这一点的成功性而言是无可争辩的，就如同 PHP 那样。

根据定义，一个全面的机器学习平台要提供算法、API、新手套件、开发和训练工具包、训练和测试数据集，以及 ML/DLC(机器学习开发生命周期)指南。Microsoft 的数据科学处理团队制作了一个合适的示例，它展示了数据科学处理如何满足软件开发生命周期的最佳实践，其中包括源代码控制、版本管理等。平台的民主化也让开发人员能够轻易地跟上研发进度；因此，其中并没有手动安装和设置。它是一个一站式服务平台，所提供的计算能力能够支持设计、清理数据、训练模型，以及将交付准备环境和生产环境部署到容器、应用程序、服务等。

这样一个民主化平台的另一方面是要提供针对 AI 所优化的硬件。由 GPU、TPU 和 FPGA 所驱动的计算是需要确保巨大的资源投入的。专门设计和架构以用于高效运行基于深度学习的计算任务的硬件，其构建和维护成本非常高。因此，民主化平台通常符合个人开发人员和初学者的最佳利益，因为他们可以利用这个已经构建好的系统。

Microsoft AI 产品组合由四个关键领域构成：代理、应用、服务和基础设施。对于代理而言，Microsoft 押注在 Cortana 上，这也是理所应当的。毕竟，它所能讲出的笑话会像下面这样的无趣冷笑话：

“小熊维尼(Winnie the Pooh)和沙皇伊凡四世(Ivan the Terrible)有什么共同之处？他们的中间名字相同。”

而且它在仅听到“Hey Cortana”的时候就会打断谈话，这些都是需要改进的。四个关键领域中的“应用”支柱包含几乎所有桌面应用程序、Office 应用，以及 .NET 或 Core 应用。作为服务的替代项，Microsoft 的 AI 生态系统包含了机器人框架(Bot Framework)、认知服务(Cognitive Services)、Cortana 智能套件，以及认知工具包(Cognitive Toolkit, CNTK)。响应命令的代理就是 Cortana，Microsoft 的机器人环境偶然地被称为机器人框架(Bot Framework)。认知服务工具包含视觉、语音、语言、知识以及搜索 API 等。机器学习工具包括 Azure ML 和认知工具包(CNTK)，还有增强现实产品 HoloLens。最后，基础设施服务包含 Azure 机器学习、Azure N Series 和 FPGA，以便为复杂机器学习和深度学习模型提供同类产品中最佳的训练支持。

Amazon 也有一款大型 AI 工具并且使用 Alexa 作为其代理。当它不会因为我们的冷笑话而偷偷摸摸并且毛骨悚然地发笑时，Alexa 的运行效果还是很棒的。Amazon Lex 是

其机器人框架，而其用于识别的认知工具则专注于基于深度学习的图像分析。Amazon Polly 是一款文本转语音引擎，而 Apache MxNet 则是驱动该认知服务的深度学习框架。

Apple 的 Siri 可能是最广为人知的数字助理之一，因为几乎所有 Apple 产品中都提供了 Siri，其中包括手机、计算机。SiriKit 是一个集成式机器人框架，Apple 的 CoreML 库借助它来帮助开发人员将机器学习模型集成到应用程序中。Apple 用于 Core ML 的方法支持用于图像分析的 Vision、自然语言处理以及 Gameplay Kit，它能很好地与 iOS 应用生态系统集成。

Facebook 提供了 Facebook Messenger 作为其代理或者说数字助理，ParlAI 是其用于对话搜索的机器人框架，FastText 和 CommAI 是其认知工具，而 PyTorch 是其机器学习和深度学习库。由于其理念是“Python 优先”，GPU-ready Tensor 库提供了具有强劲 GPU 加速的 Tensor 计算(就像 numpy 一样)，还提供了在基于磁带的自动梯度系统之上构建的深度神经网络。

IBM 对于 Watson 的生态系统进行了大量投资，并且使用它作为其认知计算平台。由于有了 Watson Virtual Agent 作为代理以及 Watson Conversation 作为机器人框架，因此 Watson 也提供了视觉、语音、语言以及 IBM 数据见解的认知 API 和服务。对于机器学习，则要使用 Watson ML 服务和 Apache System ML。其语言理解框架 Alchemy 已经不再提供服务支持，以便让位于 Watson Discovery 或 Watson Natural Language Understanding。

Google 发布了智能且无处不在的 Google Assistant 作为其代理，而其机器人框架则部分基于 API.ai 的技术，这是 Google 于 2016 年 9 月收购的一家公司。在认知领域，Google 提供了一个云端视觉 API、视频智能、语音 API、自然语言、知识图谱、自定义搜索以及 ML 高级解决方案实验室。对于机器学习，其工具构成包括云端 ML 引擎以及优秀的 ML 库 TensorFlow，现在它已等同于深度学习的代名词。TensorFlow 于 2015 年开源，其计算使用了数据流图形以便实现可扩展的机器学习。TensorFlow 拥有超过 92 000 的 GitHub 关注星级以及超过 60 000 个分支，它是整个生态系统中最活跃的机器学习和深度学习库之一。在增强现实使用场景方面，Google 使用了 Google glass@work 和 Google Day Dream/Tango。

1.2 为人工智能构建业务用例

人工智能和机器学习正帮助企业获得竞争优势，因为它们能帮助企业提升运营效率，触及新的细分客户，并且帮助企业从市场中脱颖而出。弗雷斯特研究公司(Forrester Research)预测，相较于 2016 年，2017 年人工智能领域的投资增加了超过 300%。IDC 预计，到 2020 年底，AI 市场将从 2016 年的 80 亿美元增长到 470 亿美元以上。根据 Porter 的框架(波特竞争分析框架)，企业中使用机器学习的两种关键方式就是传感和预测。传

感指的是从传感器接收大量的数据并且学习从中识别出有用的信息。应用程序和分析平台使用大量数据来预测未来行为。多亏了自动化和数字化，大多数行业中所产生的数据量都有了显著增长，而这些数据可以被转变为一种竞争优势，并可促成战略预测目标的达成。

机器学习算法被用于预测制造业中的维护需求、零售业中的库存需求、医疗健康领域中的患病风险、金融行业中供应商和购买者的潜在信用风险、交通领域中的交通模式，以及能源部门的能源使用情况。营销团队正使用机器学习来优化促销、赔偿和折扣；预测来自所有渠道的购买行为倾向；向顾客进行个性化推荐；还会预测长期顾客的忠诚度。在医疗健康领域，医护正使用计算机视觉和机器学习来诊断疾病。自然语言处理(Natural Language Processing, NLP)被用于从医学报告和文献中收集见解。基于上述概要介绍，大家是否清楚了人工智能和机器学习所带来的真实价值？接下来介绍一些使用案例。

1.2.1 自然语言理解和生成

从计算机数据中生成文本的业务用例，其价值非常高并且对于盈利是有帮助的，其中包括挖掘客户服务的上下文，生成新闻文章的摘要，以及生成财务报告。同时理解结构化和非结构化源中语句的结构与含义、情绪和意图对于现代数字化企业而言至关重要。所有主流 AI 供应商(Facebook、Microsoft、IBM Watson、Google)都提供了类似的自然语言理解和处理能力作为其云端工具的一部分。业界领先的内部部署 NLU/NLP 库包括 Stanford CoreNLP、NLTK、Genism、SpaCY 以及 textblob。

1.2.2 语音识别

通过 AI 将一些客户电话转写成文字以便执行查找(搜索)，这样的做法会比客户服务专员人工处理的速度和效率高上好几个数量级，然后将转写出的文字用作对于员工的建议以形成闭环——这听上去是不是一个合适的用例？将说出的话转换成意图和可搜索查询的做法是一个巨大的商业机会，所有主流认知数字助理工具都将其作为构成部分，这不仅能识别语音，还能生成相应的响应。

1.2.3 认知数字助理

机器人或者说虚拟代理无处不在，并且可以将其作为分析 AI 过度宣传对于技术具有不良效果的绝佳研究案例(一个重要的证据就是 Microsoft 推出的 Tay, 它最终变成了散布种族主义言论的机器人)。认知数字助理目前的用途包括客户支持、推荐系统以及智能家居管理。可用的工具包括 Amazon Alexa、Apple Siri、Google Now、IPsoft Amelia 和 Microsoft Cortana 等。

1.2.4 非结构化文本分析

分析来自不同信息孤岛中的文本是大多数组织都在应对的一项由来已久的挑战。随着时间的推移，跨组织的数据共享会变得支离破碎，而包含大量重要见解和可能的可操作数据的 PDF、Word 文档、Excel 电子表格、文本文档、wiki 以及其他数据源则会出现丢失的情况。使用非结构化文本分析技术，包括命名实体识别、链接、自动化分类和本体构建、索引、词干提取和知识图谱构造，以及与其他数据源的临时性或基于上下文的相互关系，就能使用一种有效的、上下文敏感搜索将丢失数据聚集在一起。这一具有语义知识图谱的“数据湖”让组织可以从数据中提炼出见解并且将见解转化为行动。

1.2.5 决策管理

在保险公司、金融机构、银行以及抵押贷款机构的业务领域中，都存在着处理金融规则的决策引擎。这些规则往往是静态的，而这些系统并不会根据市场情况、人口统计数据以及其他因素的变化进行学习或者调整。机器学习有助于让这些决策管理系统变得更具动态性且更加高效，这是通过引入判断逻辑来实现的，这些逻辑可能包括某个人是否可以获得贷款，某个人是否可以得到较低利率的抵押贷款，是否可以让特定的程序完成，是否可以采购一款医疗设备，或者是否可以请求货品补给。

1.2.6 机器人流程自动化

就像传统的决策管理一样，流程自动化通常是通过批处理任务、脚本以及人工操作来支持后端流程而实现的。有了机器人流程自动化和认知型 RPA，当系统通过观察人类操作员并且随后将其操作转换成脚本以便在实践中执行的时候，高效的业务流程自动化就实现了。尽管这并非主流 AI 和机器学习供应商所提供的直接产品，但机器学习平台中的几个组成部分可以被用于业界领先的工具集中，如 Automation Anywhere、Blue Prism、UiPath 和 Sales Force。

AI 以及未来的工作是一个很大的主题，无法在几页纸上进行全面介绍。本书内容将介绍各种各样的业务用例。请查阅本章的参考文献以便了解更多与企业中 AI 应用有关的内容，以及机器学习如何影响不同的行业领域和相应的机会。

1.3 机器学习的五大流派

要理解人工智能技术的状态，重要的是要熟悉其历史以及当时其生态系统的演化。这很快会让人感到费时、费力和乏味，不过幸运的是，在计算机科学研究者和华盛顿大学教授 Pedro Domingos 的著作 *The Master Algorithm* 一书中，他通过将人工智能划分为

五大不同类别而简练地对其进行了概括(见表 1-1)。

表 1-1 五大不同类别

流派	关注点	所受影响
符号主义	填补现有知识中的空白	逻辑学、哲学
联结主义	模拟大脑	神经科学
进化主义	模拟进化	生物学
贝叶斯学派	系统性降低不确定性	统计学
类推学派	关注旧项和新项之间的相似性	心理学

符号主义方法就是基于规则的系统，或者是那些相信反推技术的人所选用的方法。逆向工作，指的是采用一些假设(前提)和结论进行反推，以便之后进行逆向工作来填补空白。

第二个流派就是进化主义。进化主义坚持使用基因算法和进化编程技术。这一流派的核心是进行平行比较，并且会将进化过程中基于基因组和 DNA 的计算理念应用到数据结构中。

贝叶斯学派，这是我想要支持的思想流派，它处理的是现实生活中的不确定性。贝叶斯方法在于，在出现更多证明或信息的时候更新一种假设的可能性。贝叶斯学派所效仿的终极算法就是概率推断。贝叶斯学派使用有向无环图来应用概率图建模技术。这一流派相信，应用先验模型并且在遇到更多数据时更新假设条件的做法是正确的。

根据 Domingos 的说法，机器学习的第四个流派是类推学派。在配备了“最近邻”这一技术以及支持向量机之后，类推学派的优势就与新奇二字关联上了。

最后是联结主义，这个流派想要对人脑进行逆向工程。这一方法涉及创建人工神经元并且将它们联结成一个神经网络。如今的现代深度学习应用程序都是基于该流派方法的，该方法源自反向传播。联结主义的现代方法就被称为“深度学习”，它尤其擅长于参数估算，并且它被恰当地应用到许多领域中，比如计算机视觉、语音识别、图片处理、机器翻译以及自然语言理解。

1.4 Microsoft 认知服务——概述

在深入研究认知服务之前，重要的是要注意，这些 API 仅是 Microsoft 所提供的 AI 平台的一部分。该 AI 平台被划分成不同的部分：AI 服务、基础设施和工具，并且每个部分都包含多个环节。本书将展示来自预先构建的 AI(比如，认知服务)的一些示例，但是也会介绍该平台的其他一些部分，其中包括对话式 AI、自定义 AI、DSVM 以及深度学习框架。

API 的作用很大，因为它们会通过经常失败和快速失败帮助我们专注于业务价值命题。它们让我们可以不再重新发明轮子，并且其目的在于让人易于使用，因此也就减缓了学习曲线。在不必要费太多心思的情况下，我们就能用几行代码完成我们自己的 RESTful 实现并且添加想要的功能。它们被整合到我们选择的编程语言和平台之中，并且有许多选项可供选择，从而确保我们能够找到适合我们应用程序的选择。此外，大部分编写良好的 API 都遵循一种编码规范，这些 API 都是由对应领域的专家所编写的，并且都会提供高质量的文档、示例代码和初学者工具包，以及在需要求助时可以借助的社区支持。

Microsoft 认知服务是作为 Microsoft AI 产品的一部分来提供的一组 API，它有助于 AI 民主化和机器学习。那些像我一样从认知服务在 2015 年刚发布就开始使用它的人已经看到了这款一飞冲天的惊人产品的成熟度。Microsoft 认知服务是 API、SDK 和服务的一个集合，它包含一组不同的 API，大约有 30 个左右。这些 API 可以被分类成五个较大的类别：视觉、语音、语言、知识和搜索。

按照常规地逐个查看这些服务会让人觉得枯燥，不过我们在某些时候需要这样做，因为有契约的存在。但是我希望读者可以从现实世界用例的角度进行思考：如今，我们要如何应对这样的场景，比如要在零售展示窗口进行情感检测，以便弄清楚某位客人在可能购买一款产品或服务方面的感兴趣程度有多大，或者将情绪分析应用于一个重点小组，或者了解讲课或谈话期间听讲者的感受。我们如何才能弄清楚技术会议中的男/女比例以及帮助母语并非英语的那些人在某个活动中拥有舒适的体验？我们是否能够让盲人阅读菜单，使用人脸识别找到走失儿童，或者使用自然语言理解来处理像分类信息这样的公共论坛中所出现的人口贩卖信息？所有这些用例以及更多看似难以实现并且耗费人力的应用程序，现在都使用机器学习 API 实现了。

正如之前所讨论的，Microsoft 认知服务可以被大体分类成五个不同领域，视觉、语音、语言、知识和搜索。还有一些 API 仍处于“实验室”阶段，其中的新功能正在持续开发和测试。在这些类别和服务逐渐满足通用可用性的过程中，它们都不稳定并且容易受到变更的影响。

计算机视觉 API 让我们可以从图片和视频中获得关于面部和情绪的信息。想要从办公室内网上传数据中过滤出原始内容？可以做到。想要对图片执行 OCR(光学字符识别)以便提取文本？已经提供了。生成标题和缩略图以及跟踪识别名人——都有了。就像 Disney 的《头脑特工队》动画电影一样，我们可以使用一个简单的 API 调用来识别生气、鄙视、厌恶、恐惧、愉快、中性、悲伤和惊讶；另外要提一句，Mindy Kaling 的配音很出彩。诸如 what-dog、howold 和 captionbot 的应用都是由 Microsoft 认知服务 API 生态系统来提供服务的。计算机视觉 API 也可以应对视频处理任务，比如视频稳定，检测和跟踪人脸，生成缩略图等。业务方可能会希望识别出视频中出现某个演员的部分，或者识别出哪些内容对于特定观众是不合适的，而视觉 API 能够对这些处理提供帮助。

1.4.1 语音

这类 API 可能不会提供私人的朗诵作品，但它们非常有助于创建可以消费和生成音频流的应用程序。这些 API 可以过滤噪声、识别说话者并且识别意图。在构建我们自己的 Siri/Cortana/Alexa/Google Now Assistant 时，还可以使用 Bing Speech API 并借助语言理解智能服务(Language Understanding Intelligent Service, LUIS)来完成语音意图识别。大家可能听说过 Josh Newlan，他住在加利福尼亚，现在 31 岁，他完成了我们所有人都梦寐以求的事情。大家一定都受够了电话铃声不厌其烦的打扰，对吧？Newlan 创造了一种方法，可在拒接会议电话的同时仍旧能够表现得参与其中。该方法同时使用 Uberi 语音识别和 IBM Watson 语音转文字技术来记录会议中的发言。当会议中提到 Newlan 的名字时，他就会播放一段预先录制好的 30 秒时长的音频，以便跟上会议节奏，然后随声附和！我们也可以使用说话者识别来获悉谁正在发言。借助自定义训练，仅使用一个小的数据集就可以让准确率变得非常高。因此，如果我们仅希望收听电台节目 *Wait Wait Don't Tell Me* 中 Paula Poundstone 所讲的笑话，则可以合理地使用这一服务。

1.4.2 语言

在“认知基础调查探究(Inquiries into the cognitive bases of surveys)”一文中，斯坦福大学的 Clark 和 Schober 提到，“人们往往有一种误解，认为语言的使用主要与词语及其含义有关。实际上并非如此。语言的使用主要与使用者及其要表达的含义有关。”认知服务语言 API 可以处理自然语言，执行文本和语言分析并且加以理解。像拼写检查、情绪、语言、主题以及关键词组提取这样的常见用例，都是这一无所不包的产品功能的一部分。

1.4.3 知识

在知识经济的环境中，允许我们利用丰富知识图谱的 API 是受到极大的追捧的。Microsoft 知识(Knowledge) API 会从网络、学术界以及其他来源处收集和提炼出知识，以便探究不同方面的知识，将实体(人、位置和事件)与相关上下文联结起来，并且提供推荐建议。

1.4.4 搜索

如今搜索无处不在，已经成为我们日常生活的一部分，在需要进行搜索时，我们会毫不犹豫地选择 Google、百度或者 Bing 进行搜索，而不会在大脑中苦苦思索。由 Bing Search 所驱动的 Microsoft 认知服务搜索 API，让我们的应用程序可以访问 Bing Web、Image、Video 和 News 搜索的数十亿个网页、图片、视频和新闻以及自动建议。

Microsoft 认知服务 API 包含来自 Microsoft 研究中心的大量人工智能和机器学习研究，以及包含通过多年来所开发和运行在生产环境中的大规模应用程序而获取的数据和结果。开发这些 API 的目的在于从各种数据源处收集信息，并且应用先进的状态算法来部署软件即服务(Software-as-a-Service, SaaS)、算法即服务，或者云端的 AI 即服务。这一快速演化的机器学习 API 和 SDK 工具包旨在对开发人员赋能，驱动和促成他们可以快速构建 AI 赋能的应用程序，这些应用程序可以利用强大的语音、语言、搜索、知识和视觉相关的功能。

1.5 人工智能的伦理规范

以电影和电视节目呈现的 AI 危害人类的内容并不少见，不过《金属脑袋》(Metalhead)中可怕的机器狗在我的认知中一直是一种特殊的存在。《金属脑袋》是电视剧《黑镜》中的一集，其拍摄背景是一个并未明确指定时间的未来，并且它毫不客气地让观众完全沉浸到黑白世界中。在人类社会原因不明地崩溃之后，Maxine Peake 试图逃离机器狗的追杀。不过实际上，如果挑选潜在雇员的机器学习模型带有性别、种族、性取向或其他任何非职位相关特征的偏见，那么这个模型就容易引起恐慌，其严重性更有甚于一条幻想出来的未来主义机器狗。

人工智能和机器学习的伦理考量已经成为一个广为人们所关注的主题，因为它会直接影响就业、经济、企业、通信、运输、媒体和技术——围绕我们社交和个人生活的方方面面都会受到影响，就目前而言，由此产生的未决问题要远多于答案。Shannon Vallor 博士是 Santa Clara 大学的哲学系教授和系主任，她就这一主题编写了一本出色的书籍，并且提出了受到关切的一些关键领域。关于 AI 的焦虑源自怀疑主义和人类安全的关切，并且焦虑的主题往往是算法不透明度和自主权、机器的责任性、训练数据集多元性的缺乏，以及人类活力和智慧这一最突出的欠缺。Vallor 博士还认为，即使是最激烈的 AI 反对者也承认，AI 在各个方面相较于我们人类都是具有优势的。这些优势包括最优性、高效性、决策速度、精确度、可靠性、可读性(信息优势)、可压缩性、可重复性以及不易受到伤害。

图 1-1 所示是 Andrew Ng 和 Elon Musk 关于这一主题一次有意思的 Twitter 交流。

总之，围绕伦理规范三个关键关切点源自于 AI 相较于人类的决策优势，以目标为导向以及人类价值观的欠缺，还有就是为了避免模型偏向性的透明度或者可解释性。认为 AI 决策优势大于人类的这一观念通常难以让人接受，因为它会唤起我们关于反乌托邦式 AI 掌控世界的暑期档大片的记忆。就经典的有轨电车难题而言，相较于在这一过程中让人参与其中以便强化决策，在何时将决策管理完全移交给无监督 AI 才是合适的呢？这个问题仍旧处于较大范围的探讨之中。



图 1-1 Elon Musk(特斯拉公司的创始人和 CEO)和 Andrew Ng(一名权威的 AI 科学家和学者)之间的 Twitter 交流

即使一个组织的目标可以被归结为最大化生产效率和/或利润，或者提升股东价值，但这个组织还是必须考虑到社会责任、治理，以及当地法和国际法，或者说至少我们希望该组织能够考虑这些问题。AI 并没有这样的内在约束。将这些职责和关切添加到一个专注于目标的 AI，对于“准备好用于生产环境”的实现而言是极其重要的。AI 到底是威胁还是救星取决于我们如何定义人工智能的伦理规范，以及是否能就其达成一致共识。AI 毁灭我们之前会三思而后行吗？——这差不多相当于我们在修建水坝会冲毁蚁丘时毫不关心蚂蚁的命运一样。

1.6 结语

在反乌托邦的语境中，AI 和机器学习通常都会被描绘成这样的场景，即自动化已经导致了人类工作的消失，而人类可以选择服侍其机器主人，承担仆人工作，或者生活在贫困煎熬之中。这并非是在危言耸听。现在的组织都严重依赖 AI 和机器学习来提供数字自动化处理，不过这并不会消灭人类，而是作为一种帮助增强和加速人类开发工作的工具。这一方式的关键好处是，允许人类在更高层次的认知面上进行工作，而机器则不仅可增强和加速所有繁重乏味的工作，还可以承担重复的手动任务，尤其是那些需要大量使用搜索和知识库的任务。使用 AI 和机器学习的目标组织都拥有同一目标和关切，那就是让人类雇员的工作变得更加容易和更加有趣，这样人类雇员就可以专注于他们最想做的事情。