Broadview
www.broadview.com.cn

深度学习

理论与实战（基础篇）

李理 / 编著

Deep Learning

Theory and Practice

深述理论，注重实战

从机器学习、神经网络到TensorFlow、PyTorch和Keras

中国工信出版集团

電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

深度学习

理论与实战（基础篇）

李理 / 编著

電子工業出版社
Publishing House of Electronics Industry
北京•BEIJING

内容简介

本书不仅包含人工智能、机器学习及深度学习的基础知识，如卷积神经网络、循环神经网络、生成对抗网络等，而且也囊括了学会使用 TensorFlow、PyTorch 和 Keras 这三个主流的深度学习框架的最小知识量；不仅有针对相关理论的深入解释，而且也有实用的技巧，包括常见的优化技巧、使用多 GPU 训练、调试程序及将模型上线到生产系统中。

本书同时兼顾理论和实战，使读者既能深入理解理论知识，又能把理论知识用于实战，因此本书每介绍完一个模型都会介绍其实现，读者阅读完一个模型的介绍之后就可以运行、阅读和修改相关代码，从而可以更加深刻地理解理论知识。

回顾人工智能几十年经历过的起起落落，希望对人工智能及深度学习感兴趣的读者通过本书的学习能够更加理性、更加全面地看待这个行业，理解人工智能尤其是深度学习的原理并应用，根据当前的技术现状合理地应用深度学习去改变人们的工作、生活和学习。

图书在版编目（CIP）数据

深度学习理论与实战. 基础篇 / 李理编著.—北京：电子工业出版社，2019.7

ISBN 978-7-121-36536-2

I. ①深…II. ①李…III. ①机器学习 IV. ①TP181

中国版本图书馆 CIP 数据核字（2019）第 092406 号

责任编辑：孙学瑛

印　　刷：三河市良远印务有限公司

装　　订：三河市良远印务有限公司

出版发行：电子工业出版社

北京市海淀区万寿路 173 信箱　　邮编：100036

开　　本：787×980　1/16　　印张：26.5　　字数：537 千字

版　　次：2019 年 7 月第 1 版

印　　次：2019 年 7 月第 1 次印刷

定　　价：109.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888，88258888。

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式：（010）51260888-819，faq@phei.com.cn。

前言

我与人工智能

1997 年夏天，我在电脑上跟一个叫“将族”的中国象棋软件（CCH）下棋时，IBM 的“深蓝”击败了国际象棋世界冠军卡斯帕罗夫。当时仅觉此事离我甚远，没想到，数年之后我竟亲手编写人工智能应用，比如酒店和机票的爬虫软件、酒店知识图谱的构建、酒店评论的分析、图片的去重鉴黄、旅游领域全文搜索引擎，相关性搜索的改进和知识图谱的构建工作。没想到，人工智能会如此受关注，我作为一位从业者，当然希望人工智能能有更大的发展，希望不同行业的人能更多地了解目前人工智能，尤其是深度学习，而不希望大家对于人工智能只从一些科普文章或网文中获取过于乐观或悲观的信息。

从一位人工智能“老兵”的角度出发，对于人工智能，我认为只有真正地理解其原理，知道目前技术的边界在哪里，并动手用它解决实际问题，才是正确认知人工智能的途径，也是我编写本书的初始动机和最终目标。

本书的特点

市面上关于深度学习的书籍很多，大多分两类：一类侧重理论，多为会议论文集，更多关注基础理论和前沿进展，这类书籍通常比较难懂，而且读完之后仍然不知道怎么动手解决问题；另一类则更关注应用，多为框架工具的介绍，偶尔提及一些理论也点到为止，在读者看来各种算法只是一个黑盒子，虽然能跑起来，但是知其然不知其所以然，不知道怎么调优，碰到问题时也不知道怎么解决。

本书希望同时兼顾理论和应用，使读者既能深入理解理论知识，又能把理论知识用于实践，因此本书每介绍完一个模型都会介绍其实现，读者阅读完一个模型的介绍之后就可以运行、阅读和修改一下相关代码，从而可以更加深刻地理解理论知识。

从我开始写下第一个字到现在，前后跨越四年，一方面是因为工作忙，写作时间都是挤出来的，另一方面是我希望写好书，不奢求立言，但总归对后来进入人工智能行业的技术工程人员能有帮助。书中每一个知识点、每一行代码，均是我学懂、做会之后写下来的，里面有很多技巧也是我从多年工作当中沉淀下来的。

适合读者

我的写作目标是让本书具有自包含性，意即初入行者只要读了这本书，就什么都能学得到，也能学得会，本书读者只要有 Python 编程基础就能踏入人工智能行业的大门，因此本书不仅有人工智能、机器学习的基础知识，而且囊括了学会 TensorFlow、PyTorch 和 Keras 三个主流的深度学习框架的最小知识量；不仅有针对很多理论的深入解释，也能学到一些实用的技巧，包括常见的优化技巧、使用多 GPU 训练、调试程序及把模型上线到生产系统中。

本书主要内容

本书共包含 8 章，每章的主要内容如下：

第 1 章介绍人工智能的发展历程和机器学习的基本概念，使用通俗的语言介绍机器学习任务的分类、常见模型、损失函数和衡量指标，最后通过一个简单的线性回归示例来加深对这些概念的理解。

第 2 章介绍全连接神经网络的基本概念和反向传播算法的详细推导过程，不使用框架完全自己实现一个多层的神经网络来识别 MNIST 的手写数字。接下来介绍基本的优化技巧，包括激活函数的选择、参数的初始化、Dropout、Batch Normalization 和 Adam 等学习率自适应算法。

第 3 章介绍卷积神经网络，使用卷积神经网络来解决 MNIST 和 CIFAR-10 数据集的识别问题，通过 CIFAR-10 的例子介绍怎么在 TensorFlow 里使用多 GPU 训练，最后介绍残差神经网络。

第 4 章介绍循环神经网络，使用它来实现姓名分类及生成莎士比亚风格的句子，接着会介绍 Seq2Seq 模型和注意力机制，使用它们来实现英语—法语、汉语—英语的机器翻译功能。

第 5 章介绍生成对抗网络，介绍对抗训练的基本原理和 DCGAN 模型，最后使用 DCGAN 来实现人脸照片的生成。

第 6 章介绍 TensorFlow，首先介绍基本概念、优化器和数据输入输出等，然后介绍全连接神经网络和卷积神经网络等常见网络结构。因为 RNN 的复杂性，我们单独使用一节来详细介绍怎么在 TensorFlow 使用 RNN、LSTM 和 GRU。接着介绍高层的 Estimator API 和 TensorBoard，以及怎么调试 TensorFlow 代码。最后介绍模型的保存和 TensorFlow Serving。

第 7 章介绍 PyTorch，通过使用不同的方法来实现三层的神经网络来重点介绍 Autograd，包括数据的加载和处理，最后是一个迁移学习的示例。

第 8 章介绍 Keras，包括卷积神经网络、残差神经网络和循环神经网络在 Keras 里的用法，最后通过简短的代码示例来演示怎么实现文本图片的分类、图片问答和视频问答。

致谢

首先要感谢我的家人尤其是我的妻子对我的支持，让我有更多的时间来完成这本书。感谢我的女儿，她给了我很多的灵感和持续学习的动力。感谢出版社的编辑团队为这本书的写作和出版付出的诸多努力。最后本书的写作过程中参考了非常多的文献、资料和开源代码，感谢他们的免费开放和授权使用。

读者服务

轻松注册成为博文视点社区用户（www.broadview.com.cn），扫码直达本书页面。

- **下载资源**：本书如提供示例代码及资源文件，均可在 下载资源 处下载。
- **提交勘误**：您对书中内容的修改意见可在 提交勘误 处提交，若被采纳，将获赠博文视点社区积分（在您购买电子书时，积分可用来抵扣相应金额）。
- **交流互动**：在页面下方 读者评论 处留下您的疑问或观点，与我们和其他读者一同学习交流。

页面入口：http://www.broadview.com.cn/36536

目录

第 1 章

人工智能的基本概念

1.1 人工智能的发展历史

1.1.1 引言

人类的进化一直与工具的制造和使用息息相关，从石器时代、铁器时代、蒸汽时代、电气时代再到信息时代，先进便捷的工具一直在改变我们的生产和生活。工具的目的是延伸和拓展人类的能力。我们跑得不快，但可以借助骑马和开车日行千里；我们跳得不高，更不会飞，却能借助飞机、火箭上天。

总体来看，工具可以分为两类：拓展人类体力的工具和拓展人类脑力的工具。在计算机发明之前，人类制造的大多数工具用于减少体力劳动，比如使用牛或者拖拉机来耕地的效率更高。当然也有少量的减少脑力劳动的工具，比如算盘，也包括文字——它可以极大地扩充人类的记忆容量。

现在很多机械的脑力劳动都可以由计算机完成，如财务软件可以用于财务核算。但是它还无法实现需要“智能”才能完成的事情，比如计算机目前无法用自然语言和人类进行日常沟通。

而人工智能的目标就是让计算机能够像人类一样“智能”地解决复杂的问题。现在的人工智能系统已经能够在围棋上战胜人类世界冠军；语音识别系统已经能在某些特定场景下达到人类的识别准确率；无人驾驶的汽车也已经在某些地方实验性地上路了。未来，人工智能会有更多的应用场景，我们的终极目标是制造和人类一样甚至超越人类智能的机器（智能体）。当然这个目标能否实现还有很多争议。

闲聊一下“Computer”这个单词的两种中文翻译：电脑和计算机，分别适用于不同的场景。比如我们会说“我新买了一台笔记本电脑”，但很少说“我新买了一台笔记本计算机”。我们会说“我是学计算机专业的”，但很少会说“我是学电脑专业的”，电脑是一个更日常的词而计算机更加专业。从翻译的角度来说，计算机似乎更加忠实原意，因为早期的计算机主要用于进行各种数学运算。不知道是谁把“Computer”翻译成电脑的，但是翻译者对于它是有很高期望的——它是电路实现的“大脑”，其实也就是期望它拥有与人类大脑一样的智能。

1.1.2 发展历史

人工智能有记载的最早探索也许可以追溯到莱布尼茨，他试图制造能够进行自动符号计算的机器，但现代意义上，“人工智能”这个术语诞生于 1956 年的达特茅斯会议。人工智能有很多的定义，它本身就是很多学科的交叉融合，不同的人关注它的不同方面，因此很难给出一个大家都认可的定义。本书更多从工程应用的角度来介绍它的基本原理和应用，尽量避免不切实际的幻想和喋喋不休的争论，更关注目前或者最近的将来如何利用这些技术解决实际的问题。

黄金时期（1956—1974 年）

这个时期是人工智能的一个黄金时期，大量的资金用于支持这个学科的研究和发展。这一时期的有影响力的研究包括通用问题求解器（General Problem Solver）、最早的聊天机器人 ELIZA。很多人都以为跟他聊天的 ELIZA 是一个真人，其实它只是简单地基于匹配模板的方式来生成回复的（现在很多市面上的聊天机器人其实也使用了类似的技术）。那个时候人们非常乐观，比如 H. A. Simon 在 1958 年断言，不出 10 年计算机将在下（国际）象棋上击败人类。他在 1965 年甚至说“二十年后计算机将可以做人类能做的所有事情”。

第一次寒冬（1974—1980 年）

但到了这个时期，那些吹过的“牛”都没有实现。因此各种批评之声涌现出来，国家（美国）也不再投入更多经费，人工智能进入第一次寒冬。这个时期也是联结主义（Connectionism）的黑暗时期。1958 年 Frank Rosenblatt 提出了感知机（Perception），被认为是最早的神经网络。

兴盛期（1980—1989 年）

这一时期的兴盛得益于专家系统的流行。联结主义的神经网络也有所发展，包括 1982 年 John Hopfield 提出了 Hopfield 网络，以及同时期发现的反向传播算法。但是主流的方法仍然基于符号主义的专家系统。

第二次寒冬（1989—1993）

之前成功的专家系统由于成本太高及其他的原因，在商业上很难获得成功，人工智能再次进入寒冬期。

稳步发展期（1993—2006 年）

这一期间人工智能的主流是机器学习。统计学习理论的发展和 SVM 等工具的流行，使得机器学习进入稳步发展的时期。

迅速发展期（2006 年至今）

这一次人工智能的发展主要是由深度学习，也就是深度神经网络带动的。20 世纪八九十年代，深度神经网络虽然通过非线性激活函数解决了理论上的异或问题，而反向传播算法也使得训练浅层的神经网络变得可能。但由于计算资源和技巧的限制，当时无法训练更深层的网络，实际的效果并不比传统的“浅度”的机器学习方法好，因此并没有太多人关注这个方向。直到 2006 年，Hinton 提出了 Deep Belief Nets（DBN），通过预训练的方法使得训练更深的神经网络变得可能。2009 年，Hinton 和 DengLi 在语音识别系统中首次使用了深度神经网络（DNN）来训练声学模型，最终系统的词错误率（Word Error Rate，WER）有了极大的降低。而让深度学习在学术界声名大噪的是 2012 年的 ILSVRC 评测。在 2012 年之前，最好的 Top5 分类错误率在 25% 以上，而 2012 年 AlexNet 在比赛中首次使用了深层的卷积神经网络，取得了 16% 的错误率。之后每年都有新的好成绩出现，2014 年是 GoogLeNet 和 VGG，而 2015 年是 ResNet 残差神经网络。目前最好系统的 Top5 分类错误率在 5% 以下了。当然，真正让更多人（尤其是中国人）了解深度学习进展的是 2016 年 Google DeepMind 开发的 AlphaGo 以 4 比 1 的成绩战胜了人类世界冠军李世石。因此人工智能进入了又一次的兴盛期，各路资本竞相投入，甚至国家层面的人工智能发展计划也相继出台，很多人认为未来国家之间的竞争很重要的部分就是人工智能的竞争，人工智能尤其是深度学习似乎一下成为“显学”。

作为行业的从业者，我当然希望人工智能能有更好的发展，但回顾人工智能几十年来经历过的起起落落，我希望大家能够更加理性地看待一个行业，因为过高的期望往往容易“捧杀”它。我觉得让更多人理解人工智能尤其是深度学习的原理并能应用是一件非常有意义的事情。因为理解其原理之后就不会神化它，也就不会有不切实际的空想，也就会根据当前的技术现状合理地使用（或不使用）这项技术。因此本书会尽量用通俗易懂的方式来介绍各种模型的基本原理并通过尽可能多的代码实例来演示它们的应用。

1.2 机器学习

作为人工智能的（最重要的）一个子领域，机器学习，研究的是如何让机器自动学习出模型（当然也可以学习出规则）来解决实际的问题。研究者大多来自统计学方向，他们用统计的方法深入研究“学习”的问题，包括研究“泛化”能力的上界等问题，并提出了著名的支持向量机模型。当然更多人并不太关心这些理论，而是关注在解决实际问题中这些模型的有效性，但总的来说基于统计方法的模型是主流方法，如逻辑回归（Logistic Regression）、朴素贝叶斯（Naive Bayes）分类器、隐马尔科夫模型（Hidden Markov Model）等。

1.2.1 机器学习的基本概念

大家可能平时都写过很多程序，写程序和机器学习的思路有很大的不同。写程序时，我们是“上帝”，我们规定计算机的每一个步骤，第一步做什么，第二步做什么，这称为算法。我们能够控制所有的情况，如果出了任何问题，肯定都是程序员的责任。而在机器学习的时候，我们只是“老师”，我们告诉学生（计算机）输入是什么，输出是什么，然后期望它能够学到我们所传授的知识。比如我们跟小孩说这是狗，那是猫，我们没有办法像“上帝”那样拿着“纳米手术刀”去操作人脑神经元的连接方式，只能不断地给小孩输入“训练数据”，然后期望他能够学会什么是猫，即使我们觉得他“学会”了识别猫，也没有办法知道他是“怎么”学会的，同样的训练过程换一个人可能就会得到完全不同的结果。

机器学习和人类的学习是类似的——给它输入训练数据，然后期望它能学会。我们会给机器建立一个模型，从数学的角度来说一个模型就是一个函数，它的输入一般是一个向量（当然可以是二维的矩阵，如图片，或者三维的张量，如视频），输出可以是有限的离散的标签，如“猫”“狗”，这类问题称为分类；而如果输出是连续的值——比如用这个模型来预测气温——那么我们就称之为回归。其实人类的很多科学活动和日常生活，都是在“学习”和“应用”模型，比如开普勒通过观测大量天文数据“归纳”出行星的运动规律。从本质上讲，智能就是从“过去”学习，然后根据“现在”来预测可能的将来并根据自己的目标选择有利于自己的行为。以前只有人类能够从数据中“学习”出规律，而人工智能的目标就是让机器拥有类似人类的学习能力。

神经网络就是试图通过计算机来模拟和借鉴人脑这个模型，除神经网络之外，机器学习领域还有各种各样的模型，它们各有特点。但不管形式怎么变化，本质都是一个函数。一个（或者更准确地说是一种）模型一般都是一种函数形式，它有一些“参数”可以改变。而学习的过程就是不断调整这些参数，使得输出（尽量）接近“正确”的答案。但是一般很难预测正确，所以会定义一个损失函数（Loss Function），可以把它理解为“错误”的程度，错得越“离谱”，损失就越大。而我们的目标就是调整参数使得损失最小。

但是在“训练”数据上调好的参数在“测试”数据上也能表现好吗？这就是模型的“泛化”能力了。就和人在学校学习一样，有的同学只会做的一模一样的题，考试时稍微改变一下就不会了，这就是因为“泛化”能力太差，学到的不是最本质的东西。所以平时会定期进行“模拟考试”，来检验学生是不是真的学会了，如果考得不好，那就打回去重新训练模型、调整参数。这在机器学习里对应的就是“Validation”阶段。到最后的考试，就是最终检验的时候了，这个试卷里的题目是不能提前让人看到的，只能拿出来用一次，否则就是作弊了。对应到机器学习里就是“Test”阶段。

这里用通俗的话描述的机器学习，主要指的是监督学习。机器学习还包括无监督学习和强化学习。前者不给答案，只给数据，让人总结规律；而后者会有答案，但是答案不是现在就告诉你。人类社会里更多的是监督学习和强化学习。强化学习是获取新知识的唯一途径，即向自然学习。我们做了一个决策，其好坏可能要很长一段时间才能显现出来。而学习出来的这些知识则通过监督的方式，由家庭和学校的教育教给下一代。

机器学习可以分为三类：监督学习、非监督学习和强化学习。

1.2.2　监督学习

根据输出是连续还是离散的，把监督学习分为回归问题和分类问题。比如预测温度，输出是一个连续的实数，即回归问题。垃圾邮件分类器（如图 1.1 所示）的输出是离散（有限的），即分类问题。分类个数是 2 的叫两类分类问题；而大于 2 的则称为多类分类问题。

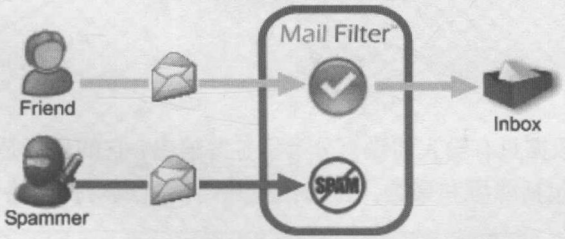

图 1.1　垃圾邮件分类器

有些问题的输入是一个序列，输出也是一个序列，如自然语言处理中的词性标注，输入是词的序列，比如“这 是 个 好 习惯”，输出是每一个词的词性，比如“PN VC M JJ NN”。我们可以把序列标注任务简化成分类问题——把每个词分类到有限个词性的集合里。但是很多词有多个词性，比如“习惯”在这里是名词，但是在句子“我习惯了这种天气”里又是动词，所以一个词的词性要参考它的上下文才能确定。我们可以简单地把当前词的前后几个词考虑进来。比如在给“习惯”进行词性标注的时候，可以考虑“我 习惯 了”这三个词的“窗口”。但要考虑多长的窗口呢？这又是个很难解决的问题。

另外对于某些序列标注任务，比如语音识别，输入和输出的长度是不一样的。在语音识别里，输入是一句话的波形文件，输出是对应的文字。如图 1.2 所示，输入和输出的长度不等。

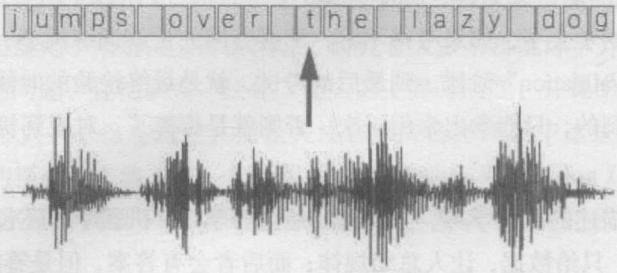

图 1.2　语音识别

除了序列，有些任务的输入和输出可能是更加复杂的结构，比如句法分析，它的输入是一个序列，而输出是更加复杂的语法树。图 1.3 是一个依存句法分析的示例。

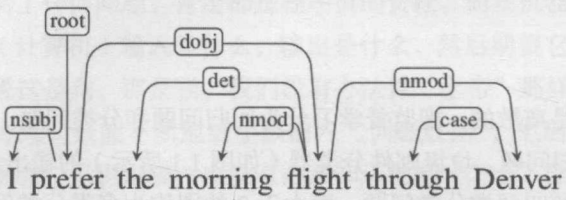

图 1.3　依存句法分析

1.2.3　非监督学习

非监督学习的训练数据只有输入而没有分类标签等输出，它的目标是根据数据的分布发现数据的规律。常见的任务包括降维和聚类，目前比较热门的生成对抗网络（GAN）就属于非监督学习。

降维的目的是把高维空间的点减低到一个低维空间上，同时尽可能多地保留信息，使得原来距离远的点在降维后也距离远而原来近的点降维后也较近。降维有很多用途，其中之一就是可视化数据，因为人眼只能理解三维以下的数据。t-SNE 是一种很常见的用于可视化的降维技术，如图 1.4 所示，我们可以把 MNIST 数据从 784 维空间降到 2 维空间。

聚类的目的是把相似的数据点放到一起，如图 1.5 所示，输入就是图中的数据点，没有任何的标注，根据数据的分布特点，聚类算法可能会把上面的数据聚类成红、绿、蓝三个类别。聚类算法通常会定义一个距离，距离越近则表示两个点越相似，让同一个聚类的点的距离较小，而不同聚类的点距离较大。

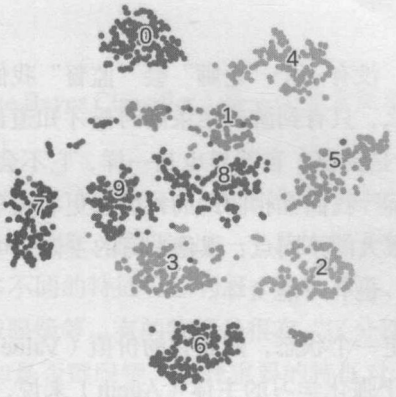

图 1.4　使用 t-SNE 对 MNIST 数据降维

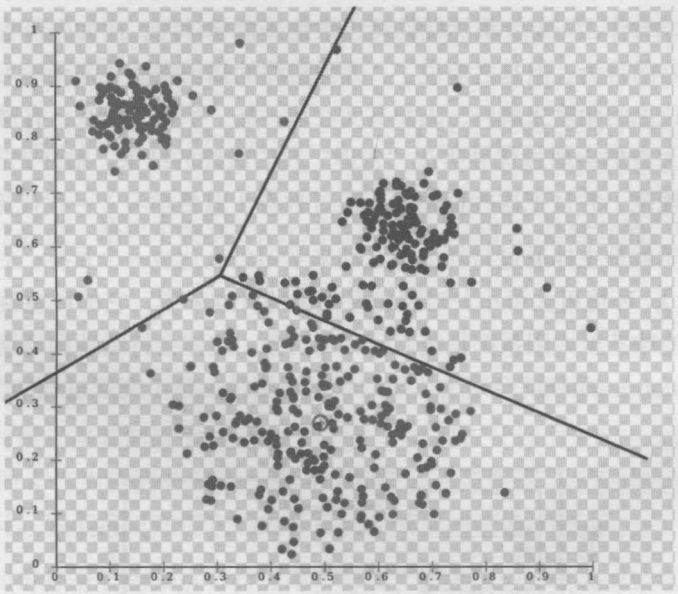

图 1.5　聚类

1.2.4　强化学习

监督学习的特点是有一个“老师”来“监督”我们，告诉我们正确的结果是什么。监督学习本质上是一种知识的传递，但不能发现新的知识。对于人类而言，真正（甚至唯一）的知识来源是实践——也就是强化学习，比如神农尝百草，最早人类并不知道哪些草能治病，但是通过尝试，就能学到新的知识。学到的这些知识通过语言文字记录下来，一代一代地流传下来，从而人类社

会作为整体能够不断进步。

强化学习和监督学习不同，没有一个“老师”会“监督”我们。比如下围棋，不会有人告诉我们当前局面最好的走法是什么，只有到游戏结束的时候才知道最终的胜负，我们需要自己复盘（学习）哪一步是好棋、哪一步是臭棋。自然界也是一样，它不会告诉我们是否应该和别人合作，但是通过优胜劣汰，最终“告诉”我们互相协助的社会会更有竞争力。和前面的监督学习、非监督学习相比，强化学习有一个很大的不同点：强化学习的主体是可以通过行为影响环境的——我们每走的一步棋都有可能变好，也有可能变坏。

它要解决的核心问题是给定一个状态，判断它的价值（Value）。价值和奖励（Reward）是强化学习最基本的两个概念。对于强化学习的主体（Agent）来说，奖励是立刻获得的，是内在的甚至与生俱来的。比如处于饥饿状态下，吃饭会有奖励。而价值是延迟的，需要计算和慎重考虑的。比如饥饿状态下去偷东西吃可以有奖励，但是从价值（价值观）的角度来看，这（可能）并不是一个好的行为。为什么不好？虽然我们可以求助于监督学习，比如先贤告诉我们这是不符合道德规范的，不是好的行为。但是之前说了，人类最终的知识来源是强化学习，先贤是从哪里知道的呢？如果从进化论的角度来解释，人类其实在玩一场“生存”游戏，有遵循道德的人群和不遵循道德的人群，大自然会通过优胜劣汰“告诉”我们最终的结果，最终先贤“学到”（其实被选择）了这些道德规范，并且把这些道德规范通过教育（监督学习）一代代流传下来。

强化学习的本质就是主体通过与环境的互动来学习怎么达成一个目标。主体交互的对象就是环境（Environment），环境可大可小，对于坐井观天的青蛙来说，它的环境就是那口小井；而对于人类来说，整个地球甚至太阳系都是我们研究的对象。主体会持续地和环境交互，根据当前的状态选择行为（Action），而环境会给主体新的状态和奖励。整个交互过程如图 1.6 所示。

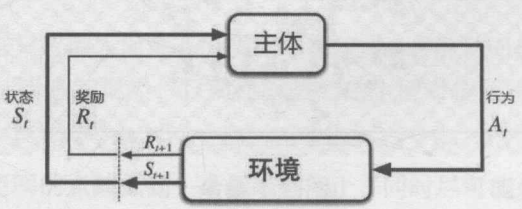

图 1.6　强化学习里主体和环境的互动

1.3　常见的监督学习模型

这里介绍一下常见的监督学习模型，希望读者通过它们能够了解机器学习的基本概念，本书主要介绍深度学习，所以这里只是简单地介绍它们的思想。