

产品经理进阶

100个案例搞懂人工智能

林中翘◎著



产品

产品经理进阶

100个案例搞懂人工智能

林中翘◎著



电子工业出版社

Publishing House of Electronics Industry

北京·BEIJING

内 容 简 介

本书根据人工智能领域产品经理的能力要求与知识体系，从原理到应用介绍人工智能的相关技术，全面阐述如何进阶为一名合格的人工智能产品经理。

本书共分为13章，其中第1~3章介绍机器学习能做什么及如何去做，第4~10章介绍7种基础算法的原理与商业化应用，第11~13章介绍深度学习在图像识别、自然语言处理与AI绘画三个方向的发展与成果。本书不局限于从数学角度推导各类机器学习算法的原理，而是配合大量案例，由浅入深地讲述什么是机器学习、机器学习如何解决问题及机器学习需要产品经理做什么。

本书能够帮助初入人工智能领域的产品经理建立对算法的理解，并将这些知识融入不同领域的业务中，发现更多的应用场景，创造更多的应用可能。

未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有，侵权必究。

图书在版编目(CIP)数据

产品经理进阶：100个案例搞懂人工智能 / 林中翘著. —北京：电子工业出版社，2019.8
ISBN 978-7-121-36498-3

I. ①产… II. ①林… III. ①机器学习—算法 IV. ①TP181

中国版本图书馆CIP数据核字(2019)第089262号

策划编辑：郑柳洁

责任编辑：梁卫红

印 刷：北京季蜂印刷有限公司

装 订：北京季蜂印刷有限公司

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：720×1000 1/16 印张：17 字数：272千字

版 次：2019年8月第1版

印 次：2019年8月第1次印刷

定 价：79.00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 zltz@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。

名家好评

谈到人工智能，相信各位读者都不会陌生，字面意思也很容易理解。在很多人眼里，人工智能是一种很先进的大数据技术，可以帮我们做很多以往计算机不能做到的事情，例如人脸识别、语音翻译、智能预测等。但是产品经理想要在自己的产品中运用人工智能技术，仅仅了解几个名词是远远不够的，还需要了解人工智能相关算法的原理及使用场景。本书作为一本专门写给产品经理的技术书，提供了大量引人入胜的案例帮助读者理解人工智能，理解算法。相信本书会对想进入人工智能领域的产品经理提供非常大的帮助。

起点学院、人人都是产品经理创始人兼 CEO，老曹（曹成明）

很多产品经理都有一些困惑：我的工作还有多久会被取代，我的存在价值有多大？产品经理转成人工智能相关的产品经理，有可能吗？有机会吗？相信本书中介绍的人工智能相关的技术原理，能帮你找到答案。

三节课联合创始人，产品经理系列课程讲师，布棉

作为一名自然语言处理方向、做过机器翻译项目的硕士，看到书中的内容颇感亲切。本书基本覆盖了常见的人工智能领域的知识点。讲述的深度恰到好处，既不会空洞无物、浅尝辄止，也不会过于深奥、成为初学者的天书。对于许多没有技术背景、欠缺算法经验的产品经理来说，阅读本书对理解人工智能大有益处。不过，作为人工智能领域的产品经理，必然需要针对用户需求和业务需要反推自己应该掌握的技能。本书可作为一本工具书，在你面对不同业务场景时，为你提供相应的解决方法。

《从点子到产品：产品经理的价值观与方法论》《产品思维》作者，
滴滴出行司机产品前负责人，刘飞

接触了很多非技术背景出身的产品经理，发现大家对技术有一个共同误区，以为学习技术需要掌握技术能力，比如写代码。实际上，从产品经理工作的角度来看，我们需要的是技术思维，即理解基本技术原理，能够以此进行产品决策。本书由浅入深全面地介绍了人工智能领域的相关技术，并通过大量案例通俗易懂地向读者介绍了机器学习算法的原理、模型的开发流程以及相关的应用，这对非技术背景出身的产品经理非常有帮助。

《产品经理必懂的技术那点事儿：成为全栈产品经理》作者，
微信公众号“唐韧”主理人，唐韧

人工智能一直都是互联网行业的热门话题，也是 B 端产品涉及的重要领域。本书通过丰富的案例，将抽象的人工智能生动地加以诠释，读起来引人入胜。这里将本书推荐给想进入人工智能领域的产品经理阅读。

《B 端产品经理必修课：从业务逻辑到产品构建全攻略》作者，李宽

最近几年，AI 已经渗透到各行各业，正以前所未有的速度和方式改变着世界。并且，与“互联网+”“区块链”等概念不同的是，AI 往往直接提升各行业的底层生产力，以最直接的方式影响工业生产率。未来，AI 很可能成为驱动世界变革的基础

技术之一，就像蒸汽机、电力和机械一样。作为产品经理，必须对这样的技术有所了解，才能将其应用到自己的产品中。而《产品经理进阶——100个案例搞懂人工智能》这本书，正好提供了一个这样的途径，为产品经理打开人工智能的大门。

《解构产品经理：互联网产品策划入门宝典》作者，
创业者，腾讯前高级产品经理，刘涵宇

很多产品经理面对扑面而来的人工智能大潮是迷茫的，他们不知道在人工智能时代，产品经理的角色到底是什么，自己到底应该如何适应和转型。包括作为广告产品经理的我也时常在想，人工智能到底会给广告行业带来什么冲击？本书在很大程度上回答了这个问题。书中通过极其丰富的案例向产品经理展示了人工智能离我们的生活有多近，它是如何运作的，以及产品经理应该如何适应。毫不夸张地说，人工智能就是新时代的蒸汽机和电，将会渗透到每一个人的生活中，而身处风口浪尖的产品经理必然要投入其中，本书或许就是一个好的开始。

资深产品经理、专栏作者，微信公众号“卫夕指北”主理人，卫夕

人工智能技术对于生活和企业的运营变得越来越重要，无论是何种领域的产品经理，都应该对人工智能技术原理有所了解，这样才能在应用层面做出合理判断和决策，采用正确的人工智能技术。作者作为一名经验丰富的人工智能产品经理，将晦涩的专业知识徐徐道来，相信本书对很多想进入人工智能领域的产品经理会有所启发和帮助。

《决胜B端：产品经理升级之路》作者，VIPKID产品总监，杨堃

前言

写作缘由

2016年，我曾看过一则新闻，讲述日本的北海道大学修建了一段不会被大雪覆盖的道路。其奥秘在于道路边上铺设了加热管，道路中设置了几个摄像头，能够通过图像识别技术检测落雪的厚度。加热管根据落雪的厚度自动调节温度，这样做能够用最低的能耗保持道路不被积雪覆盖。

这件事带给我很大的触动：原来人工智能还能这么玩，做这么酷的事情！人工智能可以帮助我们解决很“大”的问题，大到证券的量化分析、交通资源的动态配置等；也可以帮助我们解决很“小”的问题，小到让我们走在路上不用受积雪的困扰。这离不开工程师的开发能力，更离不开产品经理发现需求、寻找解决问题的能力。很荣幸我能成为其中一员，能为人工智能的发展做出一点贡献。

最初，我刚开始接触人工智能的时候，学习算法的过程非常痛苦和艰辛，主要原因在于，国内外所有人工智能相关的教材几乎都是面向专业技术人员编写的，通篇都是公式的推导与计算，很少有老师着眼于原理和场景的讲解。这对于非专业出身的产品经理来说非常不友好，晦涩难懂的公式实在是难以消化，只能花大量的时间研究资

料，慢慢理解。2018年，在电子工业出版社策划编辑郑柳洁的盛情邀请下，我萌生了为产品经理写一本算法入门书的想法。在写作过程中，我一直在思考采用什么形式才能将算法的本质讲得通俗易懂，减轻读者学习的压力。最终决定以案例讲原理，用生动的比喻代替枯燥的公式方式，让产品经理更容易接受、更容易理解。这也是本书名字的由来。

如果你想成为人工智能领域的产品经理，但又不懂技术、不懂算法，那么本书能够让你对人工智能的算法与应用有新的认知和理解，不会再觉得人工智能是一个高不可攀、遥不可及的领域。相反，人工智能是普通人也可以理解、学习和实现的，没有技术背景的产品经理也能通过学习此书，成为一名优秀的人工智能产品经理。

阅读建议

本书旨在帮助想要进入人工智能领域的产品经理掌握常见的机器学习、深度学习技术，了解不同技术的应用方式与场景，同时掌握正确的工作方法，在整个产品研发周期中体现出人工智能产品经理的价值。全书共分13章，从数据如何处理开始，到模型调优，再到算法的原理与商业化应用，由浅入深，探索人工智能的奥秘。建议读者按照章节顺序阅读，以便对机器学习有系统的认识。

第1章主要介绍什么是机器学习，以及哪些问题适合用机器学习来解决。当你机器学习有一个初步认识后，我们再具体学习机器学习有哪些步骤，如何选择模型，以及机器学习可以分为哪些类别。

第2章主要介绍数据预处理的各种方法。在实际的项目中，最初拿到手的原始数据总是存在各种各样的问题，为了让模型更好地学习数据中的规律，我们采用数据预处理的方法对原始数据进行加工。

第3章主要介绍数据的探索方法与模型的评价指标。通过对数据本质、可视化方式及模型指标三方面的探索，让产品经理对数据有更深刻的认识与理解。掌握数据的基本概念可以让我们在收集数据及进行数据预处理时更有针对性，知道哪些是不符合要求的数据，哪些是有价值的。

第4~8章主要介绍五大机器学习基础算法——回归分析、决策树、朴素贝叶斯、

神经网络和支持向量机，它们的基本原理、应用场景，以及在模型开发的过程中产品经理如何有效地解决问题。

第9~10章主要介绍集成与降维算法，这两类比较特殊的机器学习算法能够有效提升机器学习的效果。

第11~13章主要介绍深度学习在图像识别、自然语言处理与AI绘画三个方向的发展与成果。深度学习已经成为计算机视觉、语音识别和许多其他领域中机器学习的主要方法，因此也是产品经理必须关注、了解的重要领域。

致谢

感谢家人给我的大力支持，特别感谢我的妻子张容焕一直以来对我的理解与鼓励。

感谢深大互联网圈、十八罗汉工作室对本书的支持，感谢王正勇对本书的指导与帮助。

感谢平安科技同仁，感谢姜凯英、王仲秋等领导对本书的支持。

感谢电子工业出版社策划编辑郑柳洁为本书出版所付出的辛苦和努力。

最后，由于作者的水平有限，书中难免存在一些错误或不准确的地方，恳请读者批评指正。如你遇到任何不解的问题，或想提出宝贵意见，请添加微信号justinaqiao与作者交流，期待能够得到你的真诚反馈。

目录

1	机器学习入门	1
1.1	什么是机器学习.....	1
1.1.1	人类学习 VS 机器学习.....	1
1.1.2	机器学习三要素.....	3
1.2	什么问题适合用机器学习方法解决.....	5
1.2.1	必备条件.....	5
1.2.2	机器学习可解决的问题.....	7
1.3	机器学习的过程.....	9
1.3.1	机器学习的三个阶段.....	9
1.3.2	模型的训练及选择.....	11
1.4	机器学习的类型.....	12
1.4.1	有监督学习.....	13
1.4.2	无监督学习.....	14
1.4.3	半监督学习.....	14
1.4.4	强化学习.....	15
1.5	产品经理的经验之谈.....	16

2	数据的准备工作	18
2.1	数据预处理.....	18
2.1.1	为什么要做数据预处理.....	18
2.1.2	数据清洗.....	20
2.1.3	数据集成.....	23
2.1.4	数据变换.....	24
2.1.5	数据归约.....	26
2.2	特征工程.....	27
2.2.1	如何进行特征工程.....	27
2.2.2	特征构建.....	27
2.2.3	特征提取.....	28
2.2.4	特征选择.....	31
2.3	产品经理的经验之谈.....	34
3	了解你手上的数据	36
3.1	你真的了解数据吗.....	36
3.1.1	机器学习的数据统计思维.....	36
3.1.2	数据集.....	37
3.1.3	数据维度.....	41
3.1.4	数据类型.....	42
3.2	让数据更直观的方法.....	43
3.2.1	直方图.....	43
3.2.2	散点图.....	44
3.3	常用的评价模型效果指标.....	45
3.3.1	混淆矩阵.....	45
3.3.2	准确率.....	46
3.3.3	精确率与召回率.....	47
3.3.4	F 值.....	49
3.3.5	ROC 曲线.....	50
3.3.6	AUC 值.....	54

3.4	产品经理的经验之谈.....	55
4	趋势预测专家：回归分析.....	57
4.1	什么是回归分析.....	57
4.2	线性回归.....	58
4.2.1	一元线性回归.....	58
4.2.2	多元线性回归.....	63
4.3	如何评价回归模型的效果.....	66
4.4	逻辑回归.....	68
4.4.1	从线性到非线性.....	68
4.4.2	引入 Sigmoid 函数.....	71
4.5	梯度下降法.....	74
4.5.1	梯度下降原理.....	74
4.5.2	梯度下降的特点.....	76
4.6	产品经理的经验之谈.....	77
5	最容易理解的分​​类算法：决策树.....	79
5.1	生活中的决策树.....	79
5.2	决策树原理.....	80
5.3	决策树实现过程.....	82
5.3.1	ID3 算法.....	83
5.3.2	决策树剪枝.....	86
5.4	ID3 算法的限制与改进.....	88
5.4.1	ID3 算法存在的问题.....	88
5.4.2	C4.5 算法的出现.....	89
5.4.3	CART 算法.....	95
5.4.4	三种树的对比.....	97
5.5	决策树的应用.....	98
5.6	产品经理的经验之谈.....	99

6	垃圾邮件克星：朴素贝叶斯算法	101
6.1	什么是朴素贝叶斯	101
6.1.1	一个流量预测的场景	101
6.1.2	朴素贝叶斯登场	102
6.2	朴素贝叶斯如何计算	103
6.2.1	理论概率与条件概率	103
6.2.2	引入贝叶斯定理	105
6.2.3	贝叶斯定理有什么用	107
6.3	朴素贝叶斯的实际应用	108
6.3.1	垃圾邮件的克星	108
6.3.2	朴素贝叶斯的实现过程	111
6.4	进一步的提升	112
6.4.1	词袋子困境	112
6.4.2	多项式模型与伯努利模型	113
6.5	产品经理的经验之谈	114
7	模拟人类思考过程：神经网络	116
7.1	最简单的神经元模型	116
7.1.1	从生物学到机器学习	116
7.1.2	神经元模型	118
7.2	感知机	121
7.2.1	基础感知机原理	121
7.2.2	感知机的限制	125
7.3	多层神经网络与误差逆传播算法	126
7.3.1	从单层到多层神经网络	126
7.3.2	巧用 BP 算法解决计算问题	128
7.4	RBF 神经网络	132
7.4.1	全连接与局部连接	132
7.4.2	改变激活函数	134
7.5	产品经理的经验之谈	136

8	求解支持向量机.....	138
8.1	线性支持向量机.....	138
8.1.1	区分咖啡豆.....	138
8.1.2	支持向量来帮忙.....	139
8.2	线性支持向量机推导过程.....	140
8.2.1	SVM 的数学定义.....	140
8.2.2	拉格朗日乘子法.....	143
8.2.3	对偶问题求解.....	146
8.2.4	SMO 算法.....	147
8.3	非线性支持向量机与核函数.....	148
8.4	软间隔支持向量机.....	150
8.5	支持向量机的不足之处.....	152
8.6	产品经理的经验之谈.....	153
9	要想模型效果好, 集成算法少不了.....	155
9.1	个体与集成.....	155
9.1.1	三个臭皮匠赛过诸葛亮.....	155
9.1.2	人多一定力量大吗.....	157
9.2	Boosting 族算法.....	158
9.2.1	Boosting 是什么.....	158
9.2.2	AdaBoost 如何增强.....	160
9.2.3	梯度下降与决策树集成.....	163
9.3	Bagging 族算法.....	166
9.3.1	Bagging 是什么.....	166
9.3.2	随机森林算法.....	168
9.4	两类集成算法的对比.....	171
9.5	产品经理的经验之谈.....	173
10	透过现象看本质, 全靠降维来帮忙.....	175
10.1	K 近邻学习法.....	175

10.1.1	“人以群分”的算法	175
10.1.2	如何实现 KNN 算法	176
10.2	从高维到低维的转换	178
10.2.1	维数过高带来的问题	178
10.2.2	什么是降维	179
10.3	主成分分析法	180
10.3.1	PCA 原理	180
10.3.2	PCA 的特点与作用	184
10.4	线性判别分析法	186
10.5	流形学习算法	189
10.6	产品经理的经验之谈	193
11	图像识别与卷积神经网络	195
11.1	图像识别的准备工作	195
11.1.1	从电影走进现实	195
11.1.2	图像的表达	196
11.1.3	图像采集与预处理	199
11.2	卷积神经网络	202
11.2.1	卷积运算	202
11.2.2	什么是卷积神经网络	205
11.3	人脸识别技术	211
11.3.1	人脸检测	211
11.3.2	人脸识别	212
11.3.3	人脸识别的效果评价方法	214
11.4	产品经理的经验之谈	215
12	自然语言处理与循环神经网络	217
12.1	自然语言处理概述	217
12.1.1	什么是自然语言处理	217
12.1.2	为什么计算机难以理解语言	219

12.2	初识循环神经网络.....	220
12.2.1	CNN 为什么不能处理文本.....	220
12.2.2	循环神经网络登场.....	222
12.2.3	RNN 的结构.....	224
12.3	RNN 的实现方式.....	228
12.3.1	引入 BPTT 求解 RNN.....	228
12.3.2	梯度消失问题.....	230
12.4	RNN 的提升.....	231
12.4.1	长期依赖问题.....	231
12.4.2	处理长序列能手——LSTM.....	232
12.5	产品经理的经验之谈.....	235
13	AI 绘画与生成对抗网络.....	237
13.1	初识生成对抗网络.....	237
13.1.1	猫和老鼠的游戏.....	237
13.1.2	生成网络是什么.....	240
13.1.3	判别检验.....	244
13.1.4	生成对抗的过程.....	244
13.2	生成对抗网络的应用.....	246
13.2.1	GAN 的特点.....	246
13.2.2	GAN 的应用场景.....	247
13.3	生成对抗网络的提升.....	249
13.3.1	强强联合的 DCGAN.....	249
13.3.2	通过 BEGAN 化繁为简.....	251
13.3.3	对 GAN 的更多期待.....	252
13.4	产品经理的经验之谈.....	253
	参考资料.....	255

1

机器学习入门

1.1 什么是机器学习

1.1.1 人类学习 VS 机器学习

自计算机问世以来，人类一直尝试赋予计算机思想，让计算机变得更智能，使它能够理解我们说的话，看懂我们的表情，还能够帮助我们处理复杂的事情。为此，一个专门的学科诞生了，即人工智能（Artificial Intelligence）。如今，人工智能已经成为计算机科学的一个重要分支，它主要研究智能的实质，并提出一种模拟人类思考的方式。该领域的研究对象包括机器人、语音识别、图像识别、自然语言处理和专家系统等。

从 20 世纪 50 年代起，人工智能的发展进入第一阶段推理期，当时的人工智能通过赋予计算机一种简单的逻辑推理能力使它变得智能。在当时，计算机已经能够证明一些简单的数学定理，但远没有达到真正智能的标准。20 世纪 70 年代，人工智能的发展进入了第二阶段知识期。在这个时期出现了大量的专家系统，很多科学家尝试将人类的知识教给计算机，这有点像中学生的“题海战术”。但人类产生的知识量巨大，计算机没有办法全部学会，因此人工智能的发展很快就遇到了瓶颈。