

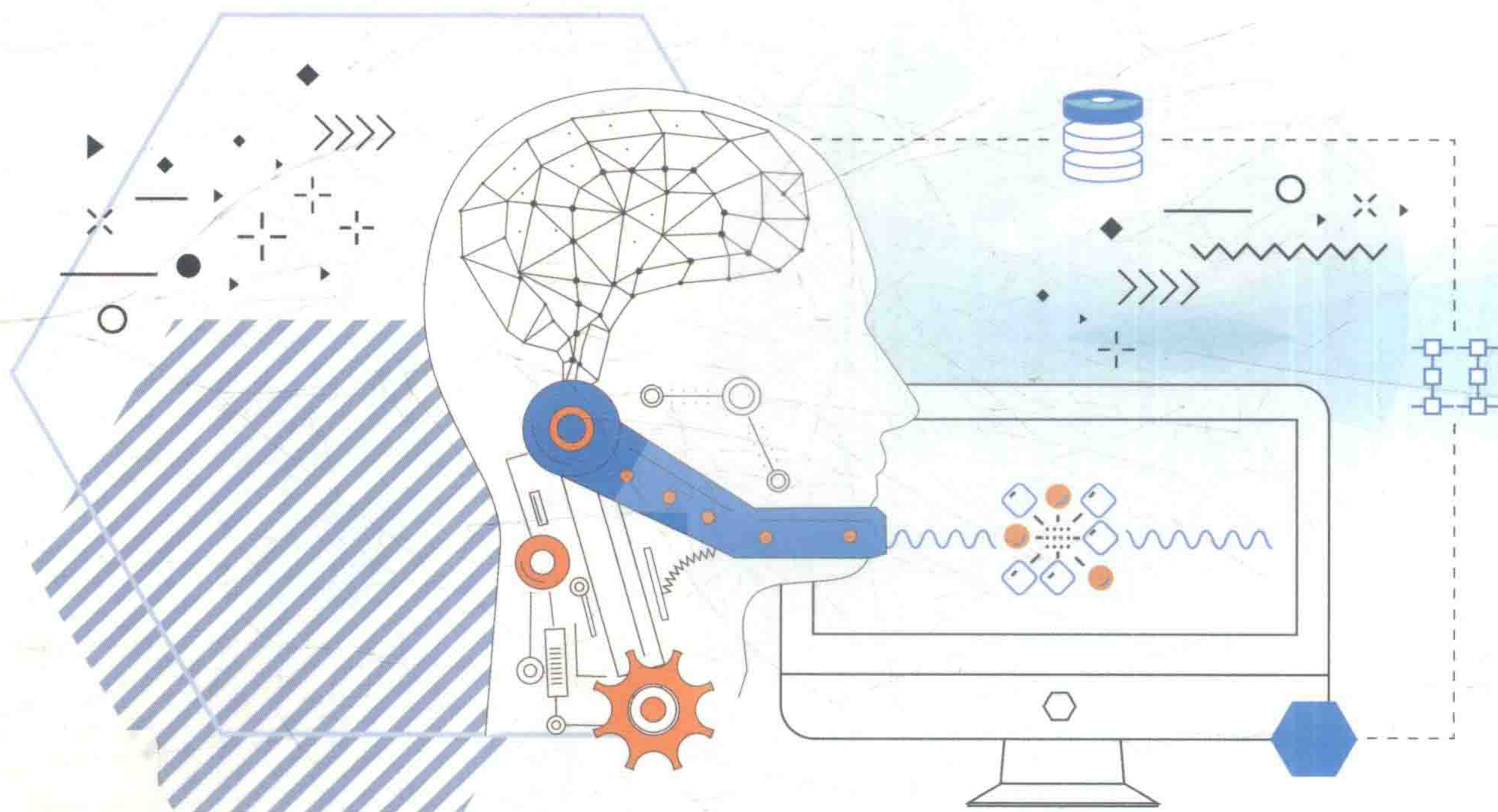
微软亚洲研究院AI工程师撰写，以实战为导向的深度学习全栈技术指南
详细讲解深度学习的知识体系、核心概念、模型、算法、工具和库，以及它们在
语音识别、聊天机器人、人脸识别、自动驾驶等领域的实践方法

Deep Learning
Core Technology, Tool and Case Study

深度学习

核心技术、工具与案例解析

高彦杰 于子叶 著



机械工业出版社
China Machine Press

Deep Learning
Core Technology, Tool and Case Study

深度学习

核心技术、工具与案例解析

高彦杰 于子叶 著



机械工业出版社
China Machine Press

图书在版编目 (CIP) 数据

深度学习：核心技术、工具与案例解析 / 高彦杰，于子叶著. —北京：机械工业出版社，2018.6

(智能系统与技术丛书)

ISBN 978-7-111-60303-0

I. 深… II. ①高… ②于… III. 学习系统 IV. TP273

中国版本图书馆 CIP 数据核字 (2018) 第 142410 号

深度学习：核心技术、工具与案例解析

出版发行：机械工业出版社（北京市西城区百万庄大街 22 号 邮政编码：100037）

责任编辑：张锡鹏

责任校对：殷虹

印刷：北京市荣盛彩色印刷有限公司

版次：2018 年 7 月第 1 版第 1 次印刷

开本：186mm×240mm 1/16

印张：17

书号：ISBN 978-7-111-60303-0

定价：69.00 元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换

客服热线：(010) 88379426 88361066

投稿热线：(010) 88379604

购书热线：(010) 68326294 88379649 68995259

读者信箱：hzit@hzbook.com

版权所有·侵权必究

封底无防伪标均为盗版

本书法律顾问：北京大成律师事务所 韩光 / 邹晓东

P R E F A C E

前 言

为什么要写这本书

随着国内各个 AI 峰会的召开，各个开源深度学习框架 meetup 的火爆举行，深度学习技术的发展正在如火如荼地进行之中，很多公司已经将深度学习技术大范围落地并且应用于生产之中。深度学习的研发人员已经从最初的通过优化算法，在公开数据集上取得不俗的成绩，转变到如今在应用的引导之下，设计优化深度学习算法并将其用于生产之中。写一本深度学习技术与实用案例类的技术图书是我持续了很久的想法。由于工作和学习较为紧张，最初只是对参与或学习过的深度学习相关案例进行了总结，但是随着时间的推移，最终我还是打算将其中通用的算法、模块流程以及应用场景抽象出来，进行适当简化并整理成书，这也算是我对深度学习的一个总结和分享。

深度学习是发源于人工神经网络的一种机器学习模型，它立足于更深的模型，具有更强的拟合能力，在计算机视觉、语音识别、自然语言处理等领域均取得了不俗的成果，并在其他应用场景中也逐步获得了傲人的成绩。深度学习已成为当下最受关注的人工智能技术，拥有着庞大的社区支持，应用场景与工具支持也日趋完善。

现在越来越多的同行已经了解并且开始应用深度学习技术。但是对于工程师和研发人员来说，市场上还缺少一本关于深度学习的实用案例类图书，很多深度学习初学者和开发人员只能参考网络上零星的技术博客或文档，这使得学习过程变得零碎且艰难，而本书正是为了解决这些问题而着意编写的。

本书希望展现给读者一个系统化的视角，秉承大道至简的主导思想，讲解深度学习的基本原理，以及如何在深度学习技术上构建复杂的人工智能应用，帮助读者开启深度学习技术应用之旅。

本书特色

从技术层面上讲，本书介绍了人工智能领域中发展势头正盛的深度学习技术和应用。深度学习作为一种机器学习算法，不仅具有模型结构的灵活性，可快速构建适用于不同领域的模型结构与算法，同时具有强大的模型拟合能力，能够在 GPU 等硬件，TensorFlow 等软件层框架的支撑下，发挥人工智能解决问题的能力，依托于深度网络的通用学习能力，快捷地构建广受欢迎的人工智能产品。

从适合读者阅读和掌握知识的角度进行结构安排，从深度学习基本概念和机制介绍入手，结合 TensorFlow 及其相关深度学习框架，构建相应的网络结构实战案例，并在最后结合不同的应用场景构建人工智能应用。本书所讲内容涵盖较为全面，不仅介绍了基本原理，如何进行深度学习，同时结合应用场景构建应用，还总结了一些实用经验。

书中一些实际操作和应用相关的章节，比较适合人工智能和机器学习开发人员参考，可以作为工作手边书；算法和优化实践方面的章节，则比较适合机器学习和算法工程师，可以分享经验，用于拓展解决问题的思路。

读者对象

- ▼ 深度学习与人工智能初学者
- ▼ 深度学习与人工智能应用开发人员
- ▼ 机器学习爱好者
- ▼ 开源软件爱好者
- ▼ 其他对深度学习与人工智能技术感兴趣的人员

如何阅读本书

本书共计 10 章，各章内容安排概括如下。

第 1 章从深度学习的概念出发，介绍深度学习的来龙去脉，阐述深度学习技术的知识图谱以及不同层面的技术概览。

第 2 章详细介绍了深度学习机制、典型模型原理与设计初衷。

第 3 章详细介绍了深度学习框架 TensorFlow，以实战加深对深度学习技术的理解，为后续使用奠定基础。

第 4 章详细介绍了 TensorFlow 的高阶特性与实战优化最佳实践。

第 5 章从实际出发，详细介绍如何通过深度学习技术构建语音识别器，采用主流深度学习语音识别模型作为实例。

第 6 章详细介绍了对话机器人领域的技术，以及如何通过深度学习技术构建对话机器人。

第 7 章详细介绍了深度学习的重要应用领域——人脸识别，以及主流的人脸识别模型与实践。

第 8 章介绍了自动驾驶领域的相关技术，以及如何通过深度学习进行车辆检测和端到端的自动驾驶解决方案。

第 9 章详细介绍了深度学习与人工智能中可视化的实例，使得研发人员能够更加直观地诊断优化和呈现结构。

第 10 章总结了深度学习训练和推断中可能会涉及的、需要注意和调节的优化点，希望他山之石可以攻玉。

如果你有一定深度学习的经验，能够理解深度学习和 TensorFlow 的相关基础知识和使用技巧，那么可以直接阅读本书第 5 ~ 10 章。如果你是一名初学者，最好从第 1 章的基础知识开始学起。

勘误和支持

由于水平有限，编写时间仓促，书中难免会出现一些错误或者不准确的地方，恳请读者批评指正。如果你有更多的宝贵意见，请反馈给我们，我们会尽量为你提供最满意的解答，欢迎读者朋友加 QQ 群（687690969）进行技术交流。

你可以通过微博 @高彦杰 gyj，公共号“戏说 AI 大数据”，邮箱“gaoyanjie55@163.com”联系到高彦杰。

你也可以通过知乎专栏“神经网络学习与 Tensorflow 实践”，邮箱“cangye@hotmail.com”联系到于子叶。

期待能够得到你的真挚反馈，在技术之路上互勉共进。

致谢

感谢机械工业出版社华章公司的编辑杨福川和张锡鹏，在近半年的时间里始终支持我们的写作，是你们的鼓励和帮助引导我们顺利完成全部书稿。

特别致谢

谨以此书献给我们最亲爱的爱人、家人、同事，以及众多热爱深度学习与人工智能技术的朋友们！

CONTENTS

目 录

前言

第 1 章 了解深度学习.....1

1.1 什么是深度学习.....1

1.1.1 深度学习能解决的问题....2

1.1.2 深度学习适用的领域.....3

1.2 深度学习的技术发展.....6

1.2.1 基础技术发展.....7

1.2.2 应用技术发展.....8

1.3 深度学习的知识点汇总.....10

1.3.1 深度学习的预备技术.....11

1.3.2 深度学习的技术构成.....12

1.3.3 深度学习向其他行业与 技术领域的延伸.....13

1.4 深度学习工具与平台介绍.....15

1.4.1 深度学习框架.....15

1.4.2 深度学习基础架构.....18

1.4.3 深度学习开发工具.....19

1.4.4 深度学习辅助工具.....20

1.4.5 深度学习云平台服务.....20

1.5 本章小结.....22

1.6 参考资料.....23

第 2 章 深度学习技术.....24

2.1 深度学习基础.....24

2.1.1 感知器.....26

2.1.2 激活函数.....27

2.1.3 输入层、隐藏层、 输出层.....29

2.1.4 前向传播与反向传播 求导.....29

2.2 CNN.....30

2.2.1 前馈网络解决图像存在的 问题.....31

2.2.2 卷积神经网络.....31

2.2.3 图像的几个不变性.....33

2.2.4 卷积.....34

2.2.5 池化.....38

2.2.6 Inception.....39

2.2.7 Flatten 层和全连接层.....39

2.2.8 跳层连接.....40

2.2.9 经典 CNN.....41

2.3	RNN	41	4.2.1	TF Layers	94
2.3.1	RNN 结构	42	4.2.2	TF Slim	95
2.3.2	基本结构与前向传播	43	4.2.3	TFLearn	96
2.3.3	BPTT	44	4.2.4	Keras	99
2.3.4	LSTM	46	4.3	Tensorboard 与问题监控	101
2.3.5	RNN 拓展	50	4.4	改善深度神经网络	103
2.4	最优化算法	52	4.5	性能优化建议	105
2.4.1	梯度下降	52	4.6	深度神经网络结构	107
2.4.2	Adam	54	4.6.1	Inception 结构	107
2.5	本章小结	56	4.6.2	ResNet 结构	108
第 3 章	TensorFlow 基础	57	4.6.3	Seq2Seq 结构	108
3.1	TensorFlow	57	4.6.4	Attention	109
3.2	获取与安装	60	4.7	本章小结	110
3.3	变量及作用域	61	第 5 章	语音识别器	111
3.4	构建计算图	66	5.1	任务分析	111
3.5	全连接网络构建	71	5.2	数据与特征分析	114
3.6	CNN 构建	78	5.2.1	语音数据库	114
3.7	RNN 构建	82	5.2.2	语音数据特征	115
3.8	多架构运行	85	5.3	主流语音识别网络结构	119
3.8.1	GPU 使用	85	5.3.1	用于语音识别的 CNN	120
3.8.2	多 CPU 使用	86	5.3.2	用于语音识别的 RNN	122
3.9	队列使用	88	5.4	CTC Loss	123
3.10	本章小结	89	5.5	文本向量化	126
第 4 章	TensorFlow 进阶	90	5.5.1	英文文本向量化	126
4.1	TensorFlow 架构与原理	90	5.5.2	中文文本向量化	129
4.2	TensorFlow 扩展	93	5.5.3	文本标签比对	130
			5.6	完整构建神经网络	131

5.6.1 构建 CNN 识别网络.....	132	7.2 Detection、Aliment 与 Identify.....	170
5.6.2 用于语音识别的 RNN.....	136	7.3 数据特征分析.....	172
5.7 数据训练.....	138	7.3.1 人脸位置和范围数据.....	173
5.8 参数调优.....	140	7.3.2 人脸关键点数据.....	174
5.9 实际数据分析.....	141	7.3.3 人脸识别数据库.....	176
5.10 本章小结.....	142	7.4 haar 分类器方式.....	178
第 6 章 对话机器人	143	7.4.1 固定特征的提取.....	178
6.1 对话机器人概述与应用领域.....	143	7.4.2 分类器.....	180
6.2 对话机器人主流技术.....	145	7.4.3 代码实现.....	182
6.2.1 基于模板的对话 机器人.....	145	7.5 神经网络方法演进.....	183
6.2.2 基于检索技术的对话 机器人.....	146	7.6 人脸识别网络构建.....	185
6.2.3 基于深度学习的对话 机器人.....	146	7.6.1 人脸检测网络 MTCNN.....	185
6.3 对话机器人的前沿与功能 扩展.....	147	7.6.2 用于人脸检测的 GoogLeNet.....	188
6.4 深度学习对话机器人原理.....	149	7.7 主流人脸识别网络差异分析.....	190
6.5 构建对话机器人.....	150	7.8 TensorFlow 搭建网络.....	191
6.5.1 Bot 界面与交互.....	150	7.9 参数调优.....	193
6.5.2 语料预处理.....	155	7.10 实战分析.....	195
6.5.3 模型构建.....	161	7.11 本章小结.....	197
6.5.4 训练流程.....	163	第 8 章 自动驾驶	198
6.5.5 在线推断.....	165	8.1 自动驾驶的介绍与应用领域.....	198
6.6 本章小结.....	166	8.1.1 自动驾驶的应用场景.....	199
第 7 章 人脸识别器	167	8.1.2 自动驾驶分级.....	199
7.1 任务分析.....	167	8.2 自动驾驶技术.....	201
		8.2.1 端到端的自动驾驶 历史.....	202

8.2.2	底层硬件支撑技术	
	平台	203
8.3	深度增强学习	204
8.4	行车检测	208
8.4.1	物体检测	208
8.4.2	YOLO 模型	209
8.4.3	车辆图像数据探索	211
8.4.4	车辆视频数据预处理	212
8.4.5	迁移学习	212
8.4.6	模型推断	213
8.4.7	绘制检测结果	213
8.5	端到端自动驾驶	216
8.5.1	英伟达 End to End	
	模型	216
8.5.2	评估指标	219
8.5.3	数据分析	219
8.5.4	读入视频，并处理	
	图像	222
8.5.5	深度学习模型构建与	
	训练	222
8.5.6	可视化结果	223
8.6	本章小结	224
8.7	参考资料	224
第 9 章 可视化实践		225
9.1	可视化发展	225
9.2	可视化过程	227
9.3	Matplotlib	228
9.4	ECharts	230

9.5	可视化实践	232
9.6	三维可视化	234
9.7	动态可视化	237
9.8	本章小结	239

第 10 章 优化实践 240

10.1	通用深度神经网络训练优化	
	建议	240
10.1.1	过拟合与欠拟合	242
10.1.2	数据增强	244
10.1.3	梯度消失	245
10.1.4	初始化权值	245
10.1.5	优化算法	246
10.1.6	超参数选择	247
10.2	深度学习系统性能优化	
	建议	250
10.2.1	输入及预处理流水线	
	优化	250
10.2.2	数据格式	251
10.2.3	编译优化	252
10.2.4	GPU 性能瓶颈诊断	252
10.2.5	CPU 瓶颈优化	254
10.2.6	模型压缩	255
10.3	工程实践建议	256
10.3.1	Model 格式转换	256
10.3.2	迁移学习	256
10.3.3	在线推断 TensorFlow	
	Serving	259
10.4	本章小结	260

了解深度学习

本章将介绍什么是深度学习以及与深度学习有关的一些重要知识，比如深度学习的一些基础概念、应用和支撑技术。深度学习的内容非常广泛，所涉及的概念也很多，其发展遵循事物发展的“螺旋上升”规律。深度学习是人工智能技术之一，随着海量数据的累积、计算力的提升以及算法的进步，以及深度学习模型的灵活性和良好的拟合能力，形成了深度学习技术突破性的飞跃和应用。

希望本章能为读者建立全面的深度学习知识体系，为后续的学习和工作打下坚实的基础。

1.1 什么是深度学习

深度学习的概念源于人工神经网络的研究，是人工智能的一个分支。包含多隐层的多层感知器就是一种深度学习结构。深度学习是通过层次结构来完成特征从形成到更加抽象的高层特征提取工作的，进而对样本进行预测。深度学习可以解决有监督学习、强化学习、无监督学习等问题。

神经网络算法是具有广泛适用性的一类算法，深度学习是人工神经网络的一个分支，具有深度网络结构的人工神经网络是深度学习最早的网络模型，深度神经网络相比于普通神经网络，其拥有更深更宽的网络结构，在此之上还有卷积层、LSTM 等新的网络组件。

深度学习所涉及的组件和网络结构的其他概念将在后续章节进行详细的介绍。

1.1.1 深度学习能解决的问题

深度学习具有解决广泛问题的能力，因为其拥有自动学习数据中的重要特征信息的能力。深度学习具备很强的非线性建模能力。众多复杂问题本质上都是高度非线性的，而深度学习实现了从输入到输出的非线性变换，这是深度学习在众多复杂问题上取得突破的重要原因之一。文本、图像、语音等问题都比较复杂，这就需要模型具有很强的非线性建模能力。

分类和回归是人工智能以及机器学习中两个非常重要的问题。

分类希望能够自动识别某种数据的类别。输入是样本，输出是样本应该属于的类别。一般情况下是这样的，首先通过已有的数据、已经标记的分类，通过训练得到可用的分类器模型，然后通过模型预测新输入数据的类别，例如针对不同类别动物的图像进行分类。举例来说，我们首先为每种动物照片指定特定的标签，例如飞机、车、鹿等，然后通过图像分类的算法建立一个分类模型，针对新的照片，分类器可以通过这个模型输出对应的标签。

如图 1-1 所示，分类器可以将图片分为三类。

回归与分类相似，但是输出需要是连续的数据，根据特定的输入向量预测其指标值。例如，我们有历史股票的时间-价格信息，价格是连续数据，通过数据训练回归模型，最终将未来的时间信息输入到训练好的回归模型，回归模型能够预测未来时间点的价格。如图 1-2 所示，可以通过深度学习模型，基于历史的收盘价，学

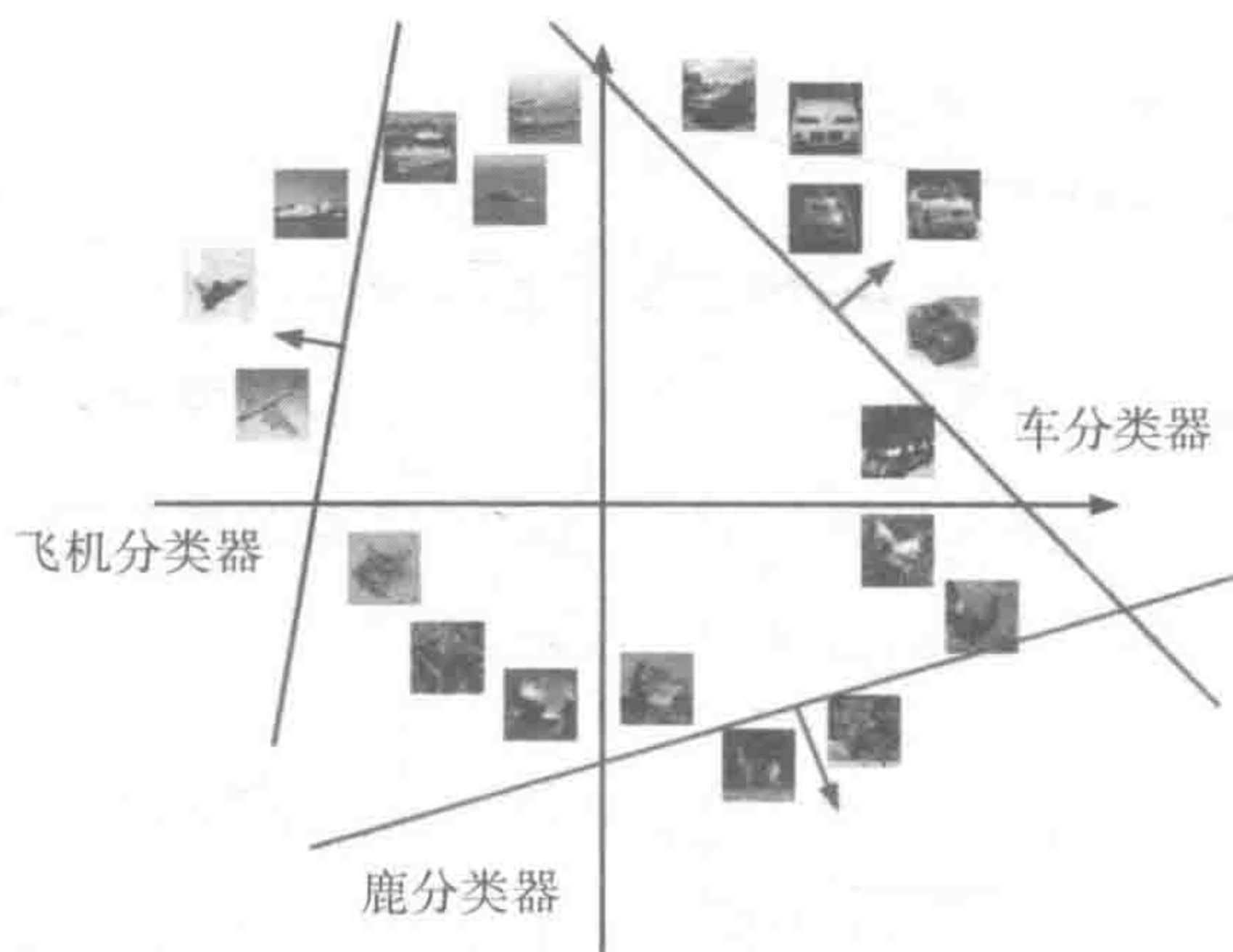


图 1-1 分类问题

习回归模型，预测未来时间点的价格。

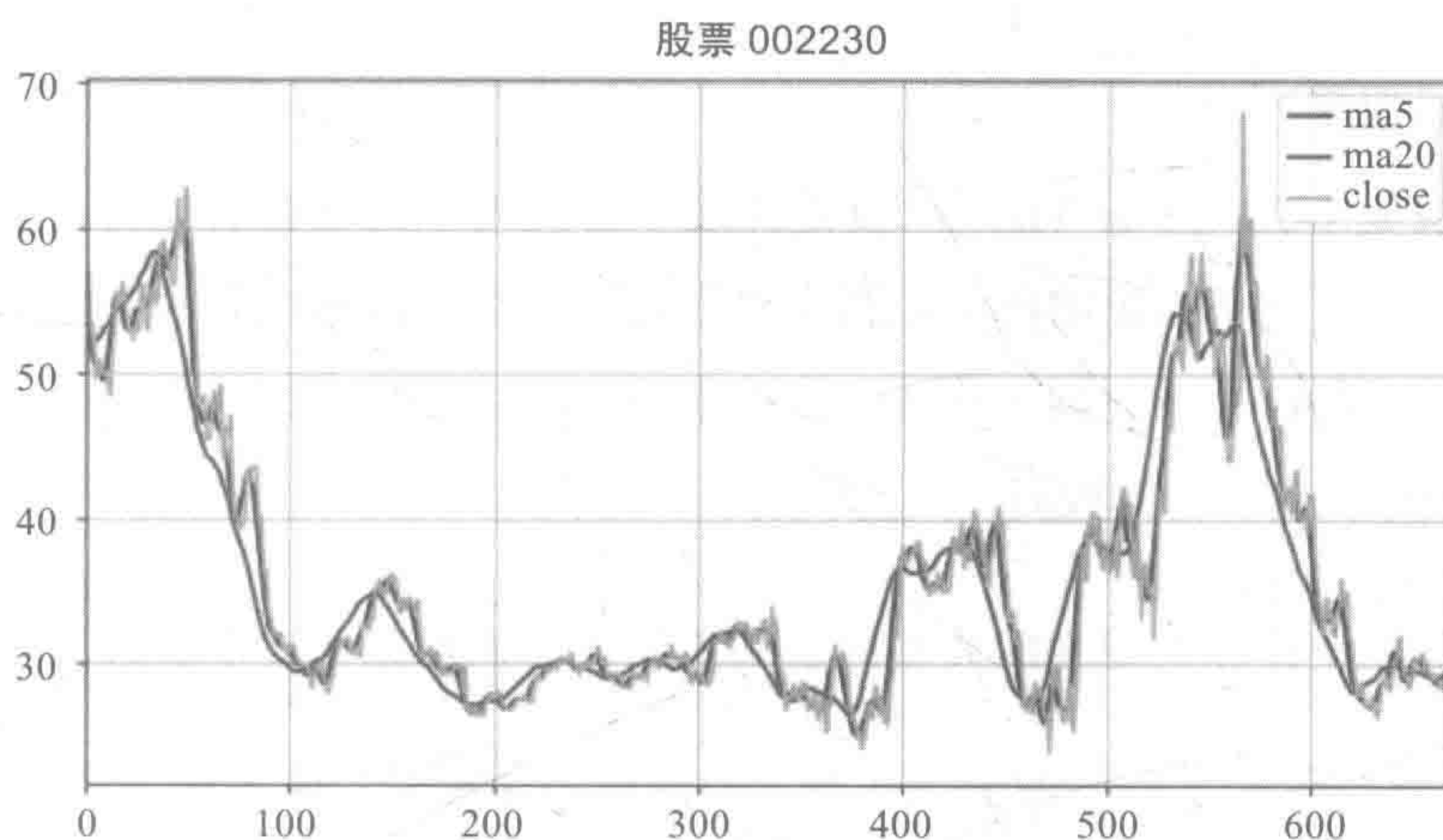


图 1-2 回归问题

深度神经网络能够灵活地对分类或回归问题进行建模和分析，这依赖于其对于特征的自动提取。例如，可以通过卷积层来表达空间相关性，通过 RNN 表达序列数据的时间连续性。另一个实例就是 Word2Vec，通过浅层神经网络可以将单词转换为向量表示。

根据问题和数据的潜在特性，研究人员设计深度学习网络结构如何从输入到输出逐步提取对预测有影响的信息，进而将不同的网络组件组织连接起来。随着大数据的增长、计算力的提升、算法模型的进化，深度神经网络模型的复杂度也变得越来越复杂，这表现为深度与广度两方面的扩展。

1.1.2 深度学习适用的领域

深度学习研究及应用的一个目标是算法及网络结构尽量能够处理各种任务，而深度学习的现状是在各个应用领域仍然需要结合领域知识和数据特性进行一定结构的设计。例如，自然语言处理任务的每一个输入特征都需要对大量的词、句等进行建模，并考虑语句中词汇之间的时间顺序。计算机视觉中的任务对每一个样本都需要处理大量的图像像素输入，并结合图像局部性等特性设计网络结构。下面是深度学习中的几个常见的适

用领域。

1. 计算机视觉

计算机视觉 (Computer Vision) 就是深度学习应用中的重要研究方向之一, 是解决如何使机器“看”这个问题的科学。因为视觉对人类以及许多动物来说毫不费力, 但是对计算机来说这个任务却充满了挑战。深度学习中有许多针对计算机视觉的细分研究和应用方向, 图像识别、物体检测、人脸识别、OCR 等。

通过以下几个计算机视觉的典型问题和应用, 我们可以快速了解计算机视觉能够解决的问题。

(1) 图像识别

图像识别问题的输入是一张图片, 输出图片中要识别的物体类别。自从 2012 年以来, CNN 和其他深度学习技术就已经占据了图像识别的主流地位。在图像识别领域有一些公开的数据集和竞赛驱动着整体技术的发展, 例如, ImageNet 是目前世界上最大的图像识别数据库, 是由美国斯坦福的计算机科学家, 模拟人类的识别系统而建立的。

(2) 物体检测

物体检测问题的输入是一张图片, 输出的是待检测物体的类别和所在位置的坐标, 通过深度学习方式可以解决。有的研究方法将问题建模为分类问题, 有的将其建模为回归问题。如图 1-3 所示, 物体检测可以识别图像中的人、狗、车辆、马等。



图 1-3 物体检测

(3) 人脸识别

人脸识别又可以细分为很多子问题, 例如, 人脸检测是将一张图片中的人脸位置识

别出来，人脸校准是将图片中人脸更细粒度的五官位置找出来，人脸识别是给定一张图片，检测数据库中与之最相似的人脸。

(4) OCR

从早期的通用扫描文档识别，到银行卡、身份证、票据等证件识别以及车牌识别都属于 OCR。

2. 语音识别

工业界和学术界掀起的深度学习浪潮在语音识别领域取得了巨大的成功。RNN 模型充分考虑了语音之间的相互关系，因此取得了更好的效果。深度学习在语音识别中的作用很大一部分表现在特征提取上，可以将其看成是一个更复杂的特征提取器。当然深度学习的作用不仅仅是特征提取，还逐渐涌现出了基于深度学习的端到端的解决方案。但这并不是说 CNN 在语音识别中无用武之地，我们一般习惯于利用 CNN 的良好训练与速度优势，而且其效果并不差。

3. 自然语言处理

深度学习在自然语言处理中的应用越来越广泛，从底层的分词、语言模型、句法分析等到高层的对话管理、知识问答、聊天、机器翻译等方面几乎全部都有深度学习模型的身影，并且取得了不错的效果。

从数据上看，经过前些年互联网的发展，很多公司都完成了海量文本数据的积累。当数据量增大时，若模型较浅，则无法对海量数据中的非线性关系进行建模，从而不能带来性能上的提升。以深度神经网络为代表的深度模型，能够随着模型复杂度的增加，对数据进行更为精准的建模，进而获得更好的效果。

从算法上看，首先，Word2Vec 的出现，使得我们可以高效地表示词，这在一定程度上缓解了语义问题，降低了输入特征的维度，从而降低了输入层的复杂度。

深度学习模型的灵活性，使得之前比较复杂的包含多流程的任务，可以使用端到端（End to End）的方法进行解决。当采用 Encoder-Decoder 架构后，模型可以将输入语言直接映射到输出语言，从而完成很多语言学问题。

4. 自动驾驶

自动驾驶也是这几年逐渐应用深度学习进行开发的一个重要领域。自动驾驶的人工智能包含了感知、决策和控制等流程和模块。感知是指通过摄像头、激光雷达等传感器的输入，进而解析出周围环境的信息，例如有哪些障碍物，障碍物的速度和距离，道路的宽度和曲率等。决策是根据感知信息判断如何进行下一步的行进规划，控制是将决策信息作用于实车。而自动驾驶中又包含了很多细分子问题，如道路与车道线的检测、前车检测、行人检测和防撞系统，以及 End to End 的自动驾驶模型等。

（1）基于规则（Rule Based）的方式

传统规则式的无人驾驶系统发展到现在，实车试验已经有十几年，研究更是有二三十年。业界和学术界主流还是基于规则系统，从车辆、传感器感知，然后进行决策、控制，最后到车辆，形成一个完整的闭环，每个流程又涉及大量的功能模块。

（2）端到端的方式

英伟达在一款汽车应用产品中，用卷积神经网络（CNN）将车前部摄像头捕捉到的原始像素图映射为汽车的方向操控命令。这种强大的端到端技术意味着，只需要人们提供少量的训练数据，系统就能学会自动驾驶技术，无论有没有车道标志线，无论是在普通公路还是高速公路上都是如此。但是深度学习在带来简便的同时也存在着难以解决的问题。

1.2 深度学习的技术发展

深度学习近年来的大范围应用和发展，除了要归功于大数据沉淀和计算能力的提升，