



好评袭来

迁移学习旨在利用已有的数据、模型和知识，通过领域相似性和“举一反三”的联想能力，把学到的通用知识适配到新的领域、场景和任务上，它使机器学习拥有更强大的泛化能力。本书作者长期和我的实验室合作，积累了丰富的科研经验，多年来辛勤地在大众媒体上普及迁移学习的相关知识。在本书中，他们保持了一贯的简明通透的写作风格，用贴近学生群体的语言，将迁移学习的发展历史、基础知识和最新进展娓娓道来。同时，本书配有用于实践的源码和数据集，增加了动手练习的环节，提高了趣味性。作为长期耕耘在迁移学习这一人工智能领域的学者和业界首本迁移学习著作（《迁移学习》）的作者，我强力推荐这本书给有志于从事迁移学习研究的同学，更快地入门和学习！

——杨强 微众银行首席人工智能官、香港科技大学讲席教授，
ACM/AAAI/IEEE Fellow

迁移学习是机器学习的一个重要研究分支，有广泛的应用价值。该书叙述简洁明了、内容丰富详实，对希望了解并应用迁移学习的读者很有帮助！

——周志华 南京大学教授，ACM/AAAI/IEEE Fellow

迁移学习对于增强训练模型的适应性具有重要意义，受到很多学者的关注。这本书深入浅出、系统性地介绍了主要的迁移学习方法，并结合多个领域的应用进行示例分析，为从事相关技术的研究人员提供了非常有益的参考。

——陶建华博士 中国科学院自动化研究所研究员，
模式识别国家重点实验室副主任

迁移学习的核心思想中国早已有之，如《周易》云：“引而伸之，触类而长之，天下之能事毕矣也”。如今，迁移学习已成为人工智能的一项核心技术，在

计算机视觉、自然语言语音处理、强化学习中得到了广泛的应用。本书语言简洁、内容丰富，相信可以启发读者举一反三、触类旁通，更好地解决手头的问题。

——秦涛博士 微软亚洲研究院首席研究经理，中国科技大学兼职教授

迁移学习是机器学习的一个重要领域。在计算机视觉，自然语言处理，语音识别，推荐系统等领域有非常广泛的应用。陈益强和王晋东两位老师通俗易懂地介绍了迁移学习的来龙去脉——不仅涵盖了基本的理论脉络、具体的方法和技术，还介绍了广泛的应用案例和未来的发展方向和前沿问题，为人工智能初学者提供了一份难得的、快速入门的学习和研究资料。

——汪军 伦敦大学学院计算机系教授

迁移学习，借用了面向对象编程的概念（模型层面的继承）、是对已训练得到的机器学习模型的高效重用，能很大程度避免资源的重复消耗，是大模型民主化的重要途径之一。本书详细介绍了迁移学习的概念和技术及最新的预训练、知识蒸馏、元学习等研究方向，内容上可谓面面俱到。除此之外，本书的一大亮点，是对“两头”的把握：一是源头，抓问题和场景，做到“师出有名”，讲清楚针对什么问题、用在哪里；二是笔头，抓代码与实践，做到“落地结果”，在实战中巩固和深化对技术的理解。相信这本书能带给读者思考与实践的双重乐趣，在算力爆炸的时代反思机器学习的高效之道！

——陈光 北京邮电大学副教授，新浪微博 @ 爱可可-爱生活



写在前面

机器学习作为人工智能领域的重要分支，在近几年取得了飞速的发展。机器学习使计算机能够从大量的训练数据和经验中学习，并将此能力应用于未知的问题和环境。迁移学习是机器学习的一种重要学习范式，旨在研究如何让已有的算法、模型、参数能够快速适用到新的问题中。

随着人工智能和机器学习的发展，迁移学习的原理、算法、模型也经历了井喷式的大发展，相关的研究工作如汗牛充栋。在海量的资料面前，这一领域的研究者、特别是初学者，难以发掘最有启发性的、本质的内容，就好比雾里看花、水中望月一般。迁移学习领域迫切需要一本能够由浅入深、由表及里阐述已有研究工作的读物，以帮助领域研究人员快速建立起这一学科的知识体系。因此，笔者在 2018 年开源了《迁移学习简明手册》，初衷便是希望用通俗易懂的内容帮助读者快速入门这一领域——这成为本书的缘起之一。

2020 年，新冠肺炎疫情打乱了每个人的工作和学习计划，也带给我们更多的思考。在此期间，适逢杨强教授《迁移学习》专著出版，笔者得以从中学习更多知识、并做了更深入的思考和总结。在出版社的邀请与帮助下，笔者启动了写书计划，写作过程几乎耗尽了笔者 2020 年绝大多数的周末和公共假期。

本书建立在笔者近几年在中国科学院大学开设的普适计算课程相关课件以及笔者前期开源的《迁移学习简明手册》的基础上，重点考虑如何从学生入门的角度循序渐进地引入迁移学习的相关概念、问题、方法和应用。更重要的是，和其他参考书着重介绍某种方法不同，本书不再侧重阐述某类特定的方法或某篇特定的论文，而是试图从学生学习的视角，归纳、总结不同类型的迁移学习方法，并结合笔者自己的理解和实践，总结成相关的文字材料。笔者希望这种“讲课”而非“学术报告”的方式能够让更多有志于迁移学习的同学更快地了解此领域，并将其应用于解决自己的问题。

当然，与国内外诸多专家学者相比，笔者深感自己能力之不足。书中如有错误和疏忽之处，恳请广大读者批评指正。

作者

2021年3月

致谢

笔者在撰写本书过程中得到了许多人的帮助，在此对他们表示感谢。

内容撰写：感谢微软亚洲研究院研究员刘畅博士协助撰写“基于因果关系的迁移学习”一节、南京大学博士生杜云涛协助撰写“迁移学习理论”和“在线迁移学习”两节、中科院计算所博士生朱勇椿协助撰写“多源迁移学习”一节。

全书修改意见：感谢微众银行首席人工智能官、香港科技大学讲席教授杨强教授、南京大学周志华教授、微软亚洲研究院首席研究经理秦涛博士、新加坡国立大学的 Research Fellow 冯文杰博士、西安电子科技大学的段然博士、大连理工大学的博士生王维，以及中科院计算所博士生秦欣、卢旺、于超辉（现任阿里巴巴达摩院算法工程师）提供的宝贵修改意见。

感谢出版社提供的专业出版意见和支持。

最后，在撰写本书的过程中，笔者得到了家人的大力鼓励和支持，在此特别表示深深的谢意。



前言

本书的编写目的是帮助迁移学习领域的初学者快速入门。本书尽可能绕开过于理论化的概念，专注介绍经验方法。除此之外，本书还配有相关的代码、数据和论文资料，最大限度地方便初学者学习。

本书共分四大部分：背景与概念、方法与技术、扩展与探索，以及应用与展望。

第一部分为背景与概念，由第 1 章到第 3 章构成。其中第 1 章为绪论，从宏观角度介绍了迁移学习的基本概念及其必要性，并且简单分析了它与已有概念的区别和联系。这一章也介绍了迁移学习的一些应用领域，目的是使读者对迁移学习有较为系统的了解。第 2 章从机器学习开始，逐步过渡到迁移学习的概念上。第 3 章介绍了迁移学习领域的基本研究问题。

第二部分为方法与技术，这是全书最重要的部分，由第 4 章到第 11 章构成。第 4 章以较为严谨的学术风格对迁移学习的基本问题进行了形式化定义，并描述了一个较为完整的迁移学习过程，以及对迁移学习理论分析的一些总结。这一章应该视为余下章节的起点。

第 5~8 章对应三大类迁移学习的基本方法：第 5 章对应样本权重迁移法，第 6 章、第 7 章分别介绍基于统计距离和几何特征的特征变换迁移方法，这两章合起来对应特征变换迁移法。由于此类方法的相关工作最为丰硕，因此我们分为两个章节讲述。第 8 章则对应基于模型的迁移，特别是在深度模型中的预训练方法。

第 9 章和第 10 章重点讲述深度迁移学习和对抗迁移学习的基本思路和方法。读者应当注意的是，深度网络中的迁移方法不应当与之前的三大类基本方法割裂开，而应该被视为三种基本方法在深度网络中的具体体现。因此，这也是为什么我们不直接谈深度方法而首先介绍三大类基本方法的原因。

第 11 章介绍了迁移学习领域若干热门研究问题和相关工作。这些从不同视角出发的问题从各个方面对经典的迁移学习场景进行了扩展，在目前仍然是热

门的研究方向。

第三部分为扩展与探索，由第 12 章到第 14 章构成。所谓扩展，指的是不局限于固定的迁移学习问题，旨在探索迁移学习新方向的一些研究成果。我们重点选择了领域泛化（第 12 章）和元学习（第 13 章）这两个研究方向进行探究和分析。第 14 章则给出了在迁移学习模型选择方面的一些代表工作。

第四部分为应用与展望，由第 15 章和第 16 章构成。第 15 章是迁移学习的应用，介绍了迁移学习在包括计算机视觉、自然语言处理、语音识别、普适计算、医疗健康等领域要解决的问题及应用的方式，向读者展示迁移学习是如何被应用到特定的任务，用以解决该应用的痛点问题的。读者将迁移学习应用于自己的任务时，可以借鉴本章所述的应用及解决方案。第 16 章介绍了几个迁移学习的前沿问题。

另外，附录部分提供了一些常用的研究资料，供初学者学习研究。

由于作者水平有限，不足和错误之处，敬请不吝批评指正。

符号表

x	变量
\mathbf{x}	向量
\mathbf{A}	矩阵
\mathbf{I}	单位阵
\mathcal{X}	输入空间
\mathcal{Y}	输出空间
\mathcal{D}	数据领域、数据集
\mathcal{N}	正态分布
\mathcal{H}	假设空间，或希尔伯特空间
$P(\cdot)$	概率密度函数
$P(\cdot \cdot)$	条件概率密度函数
$k(\cdot, \cdot)$	核函数
$\mathbb{E}_{\cdot \sim \mathcal{D}}[f(\cdot)]$	函数 $f(\cdot)$ 在数据集 \mathcal{D} 上的期望
$\ell(\cdot, \cdot)$	损失函数
$\mathbb{I}(\cdot)$	指示函数，当 \cdot 为真时取值为 1，否则为 0
$\{\dots\}$	集合
\mathbf{A}^T	矩阵 \mathbf{A} 的转置
$\text{tr}(\mathbf{A})$	矩阵 \mathbf{A} 的迹
$\max f(\cdot), \min f(\cdot)$	函数 $f(\cdot)$ 的最大值、最小值
$\arg \max f(\cdot), \arg \min f(\cdot)$	函数 $f(\cdot)$ 取最大（最小）值时对应参数的取值
$\ \cdot\ _p$	p -范式 (Norm)
$\sum_{i=1}^n i$	求和

术语表

简称	英文全称	中文全称
AutoML	Automated Machine Learning	自动机器学习
BN	Batch Normalization	批归一化
CNN	Convolutional Neural Networks	卷积神经网络
CV	Computer Vision	计算机视觉
DA	Domain Adaptation	领域自适应
DG	Domain Generalization	领域泛化
EM	Expectation Maximization	期望最大化算法
ERM	Empirical Risk Minimization	经验风险最小化
GAN	Generative Adversarial Networks	生成对抗网络
KD	Knowledge Distillation	知识蒸馏
MAP	Maximum A Posteriori	最大后验估计
ML	Machine Learning	机器学习
MLE	Maximum Likelihood Estimation	最大似然估计
MLP	Multi-layer Perceptron	多层感知机
MMD	Maximum Mean Discrepancy	最大均值差异
NLP	Natural Language Processing	自然语言处理
NMT	Neural Machine Translation	神经机器翻译
NT	Negative Transfer	负迁移
OT	Optimal Transport	最优传输
PTM	Pre-trained Model	预训练模型
RKHS	Reproducing Kernel Hilbert Space	可再生核希尔伯特空间
RL	Reinforcement Learning	强化学习
RNN	Recurrent Neural Networks	循环神经网络
SGD	Stochastic Gradient Descent	随机梯度下降
SRM	Structural Risk Minimization	结构风险最小化
SVM	Support Vector Machines	支持向量机
TL	Transfer Learning	迁移学习
TTS	Text-to-Speech	语音合成、文字转语音



目录

第一部分 背景与概念

第 1 章 绪论	3
1.1 迁移学习	3
1.2 相关研究领域	6
1.3 迁移学习的必要性	8
1.3.1 大数据与少标注之间的矛盾	8
1.3.2 大数据与弱计算能力的矛盾	8
1.3.3 有限数据与模型泛化能力的矛盾	9
1.3.4 普适化模型与个性化需求的矛盾	9
1.3.5 特定应用的需求	10
1.4 迁移学习的研究领域	11
1.4.1 按特征空间分类	12
1.4.2 按目标域有无标签分类	12
1.4.3 按学习方法分类	12
1.4.4 按离线与在线形式分类	13
1.5 迁移学习的应用	14
1.5.1 计算机视觉	14
1.5.2 自然语言处理	15
1.5.3 普适计算与人机交互	16
1.5.4 医疗健康	17
1.6 学术会议和工业界中的迁移学习	18

第 2 章 从机器学习到迁移学习	21
2.1 机器学习及基本概念	21
2.2 结构风险最小化	22
2.3 数据的概率分布	23
2.4 概念与符号	25
2.5 迁移学习的问题定义	26
第 3 章 迁移学习基本问题	29
3.1 何处迁移	30
3.2 何时迁移	32
3.3 如何迁移	32
3.4 失败的迁移：负迁移	33
3.5 完整的迁移学习过程	35

第二部分 方法与技术

第 4 章 迁移学习方法总览	39
4.1 迁移学习总体思路	39
4.2 分布差异的度量	40
4.2.1 百花齐放的迁移学习分布度量	41
4.2.2 分布差异的统一表征	42
4.2.3 分布自适应因子的计算	44
4.3 迁移学习统一表征	45
4.3.1 样本权重迁移法	46
4.3.2 特征变换迁移法	47
4.3.3 模型预训练迁移法	48
4.3.4 小结	48
4.4 上手实践	48
4.4.1 数据准备	49

4.4.2	基准模型构建: KNN	51
4.5	迁移学习理论	53
4.5.1	概念与符号	54
4.5.2	基于 \mathcal{H} -divergence 的理论分析	54
4.5.3	基于 $\mathcal{H}\Delta\mathcal{H}$ -distance 的理论分析	55
4.5.4	基于差异距离的理论分析	57
4.5.5	结合标签函数差异的理论分析	58
第 5 章	样本权重迁移法	59
5.1	问题定义	59
5.1.1	样本权重迁移法的可行性分析	59
5.1.2	形式化定义	60
5.2	基于样本选择的方法	61
5.2.1	基于非强化学习的样本选择法	62
5.2.2	基于强化学习的样本选择法	63
5.3	基于权重自适应的方法	64
5.4	上手实践	66
5.5	小结	68
第 6 章	统计特征变换迁移法	69
6.1	问题定义	69
6.2	最大均值差异法	70
6.2.1	基本概念	70
6.2.2	基于最大均值差异的迁移方法	72
6.2.3	求解与计算	75
6.2.4	应用与扩展	76
6.3	度量学习法	78
6.3.1	从预定义的距离到可学习的距离	78
6.3.2	度量学习及其形式化	79

6.3.3 基于度量学习的迁移学习	80
6.4 上手实践	81
6.4.1 算法精炼	81
6.4.2 编写代码	82
6.5 小结	84
第 7 章 几何特征变换迁移法	85
7.1 问题定义	85
7.2 子空间变换法	86
7.3 流形学习法	87
7.3.1 流形学习	87
7.3.2 基于流形学习的迁移学习方法	88
7.4 最优传输法	91
7.4.1 最优传输	91
7.4.2 基于最优传输法的迁移学习方法	92
7.5 上手实践	94
7.6 小结	97
第 8 章 预训练方法	99
8.1 深度网络的可迁移性	99
8.2 预训练-微调	102
8.3 预训练方法的有效性分析	105
8.4 自适应的预训练方法	106
8.5 重新思考预训练模型的使用	108
8.6 上手实践	110
8.7 小结	113
第 9 章 深度迁移学习	115
9.1 总体思路	116
9.2 深度迁移学习的网络结构	117

9.2.1	单流结构	118
9.2.2	双流结构	118
9.3	数据分布自适应的深度迁移学习方法	120
9.3.1	边缘分布自适应	120
9.3.2	条件、联合与动态分布自适应	121
9.4	结构自适应的深度迁移学习方法	122
9.4.1	批归一化	123
9.4.2	批归一化用于迁移学习	123
9.4.3	基于多表示学习的迁移网络结构	124
9.5	知识蒸馏	125
9.6	上手实践	127
9.6.1	网络结构	127
9.6.2	损失	129
9.6.3	训练	131
9.6.4	测试	132
9.7	小结	133
第 10 章 对抗迁移学习		135
10.1	生成对抗网络	135
10.2	对抗迁移学习基本思路	136
10.3	数据分布自适应的对抗迁移方法	137
10.4	基于信息解耦的对抗迁移方法	140
10.5	基于数据生成的对抗迁移方法	141
10.6	上手实践	142
10.6.1	领域判别器	143
10.6.2	分布差异计算	143
10.6.3	梯度反转层	144
10.7	小结	145

第 11 章 迁移学习热门研究问题	147
11.1 类别不均衡的迁移学习	148
11.2 多源迁移学习	150
11.3 开放集迁移学习	153
11.4 时间序列的迁移学习	154
11.5 联邦迁移学习	158
11.5.1 联邦学习	158
11.5.2 联邦迁移学习	160
11.6 基于因果关系的迁移学习	161
11.6.1 什么是因果关系	161
11.6.2 因果关系与迁移学习	163
11.7 自动迁移学习	168
11.8 在线迁移学习	171

第三部分 扩展与探索

第 12 章 领域泛化	177
12.1 领域泛化问题	177
12.1.1 背景	177
12.1.2 问题定义	179
12.1.3 常用方法	180
12.2 基于数据分布自适应的方法	181
12.2.1 领域无关成分分析 DICA	181
12.2.2 深度数据分布自适应	183
12.3 基于解耦的方法	184
12.4 基于集成模型的方法	186
12.5 基于数据生成的方法	187
12.5.1 领域随机法	187

12.5.2	对抗数据生成	188
12.6	基于元学习的方法	190
12.7	小结	191
第 13 章	元学习	193
13.1	元学习简介	193
13.1.1	问题背景	193
13.1.2	元学习	194
13.2	基于模型的元学习方法	196
13.3	基于度量的元学习方法	198
13.4	基于优化的元学习方法	199
13.5	元学习的应用与挑战	201
13.5.1	应用	201
13.5.2	现存的挑战	202
13.6	小结	202
第 14 章	迁移学习模型选择	205
14.1	模型选择	205
14.2	基于密度估计的模型选择	206
14.3	迁移交叉验证	207
14.4	小结	208

第四部分 应用与展望

第 15 章	迁移学习的应用	211
15.1	计算机视觉	212
15.2	自然语言处理	214
15.3	语音识别与合成	216
15.4	普适计算与人机交互	218

15.5	医疗健康领域	220
15.6	其他应用	223
15.7	小结	225
第 16 章	迁移学习前沿	227
16.1	融合人类经验的迁移	227
16.2	迁移强化学习	228
16.3	迁移学习的可解释性	228
16.4	迁移学习系统	229
附录 A		231
A.1	常用度量准则	231
A.1.1	常见的几种距离	231
A.1.2	余弦相似度	232
A.1.3	互信息	232
A.1.4	相关系数	232
A.1.5	KL 散度与 JS 距离	233
A.1.6	最大均值差异 MMD	233
A.1.7	Principal Angle	234
A.1.8	\mathcal{A} -distance	234
A.1.9	希尔伯特-施密特独立性系数	234
A.1.10	Wasserstein Distance	234
A.2	迁移学习常用数据集	235
A.2.1	手写体识别图像数据集	235
A.2.2	对象识别数据集	236
A.2.3	图像分类数据集	237
A.2.4	通用文本分类数据集	237
A.2.5	行为识别公开数据集	238
A.3	本书相关资源	238
	参考文献	241

