

《解析深度学习：语音识别实践》

图书基本信息

书名：《解析深度学习：语音识别实践》

13位ISBN编号：978712128796X

出版时间：2016-6

作者：俞栋,邓力

页数：336

译者：俞凯,钱彦旻

版权说明：本站所提供下载的PDF图书仅提供预览和简介以及在线试读，请支持正版图书。

更多资源请访问：www.tushu000.com

《解析深度学习：语音识别实践》

内容概要

AlphaGo与李世石的围棋大战激发了人们对人工智能是非的诸多争论。人工智能背后的工作原理深度学习跳入大众视野。AlphaGo的大获全胜一定程度展示了深度学习在应用领域的成功，而语音识别正是深度学习取得显著成功的应用领域之一。

本书是首次以深度学习为主线介绍语音识别应用的书籍，对读者了解语音识别技术及其发展历程有重要的参考价值。

本书作者俞栋、邓力均是该领域的著名专家，他们是深度学习在应用领域取得突破性进展的推动者与实践者，他们在书中分享的研究成果一定程度上代表了本领域最新的研究进展；译者俞凯、钱彦旻也是本领域的资深专家，并有众多实践成果。对于从事此领域研究的读者来说，本书无疑有重要的参考价值。

《解析深度学习：语音识别实践》是首部介绍语音识别中深度学习技术细节的专著。全书首先概要介绍了传统语音识别理论和经典的深度神经网络核心算法。接着全面而深入地介绍了深度学习在语音识别中的应用，包括“深度神经网络-隐马尔可夫混合模型”的训练和优化，特征表示学习、模型融合、自适应，以及以循环神经网络为代表的若干先进深度学习技术。

《解析深度学习：语音识别实践》适合有一定机器学习或语音识别基础的学生、研究者或从业者阅读，所有的算法及技术细节都提供了详尽的参考文献，给出了深度学习在语音识别中应用的全景。

作者简介

俞栋

1998年加入微软公司，现任微软研究院首席研究员、浙江大学兼职教授和中科大客座教授。他是语音识别和深度学习方向的资深专家，出版了两本专著，发表了150多篇论文，是近60项专利的发明人及有广泛影响力的深度学习开源软件CNTK的发起人和主要作者之一。他在基于深度学习的语音识别技术上的工作带来了语音识别研究方向的转变，极大地推动了语音识别领域的发展，并获得2013年IEEE信号处理协会最佳论文奖。俞栋博士现担任IEEE语音语言处理专业委员会委员，曾担任IEEE/ACM音频、语音及语言处理汇刊、IEEE信号处理杂志等期刊的编委。

邓力

世界著名人工智能、机器学习和语音语言信号处理专家，现任微软首席人工智能科学家和深度学习技术中心研究经理。他在美国威斯康星大学先后获硕士和博士学位，然后在加拿大滑铁卢大学任教获得终身正教授。其间，他还任麻省理工学院研究职位。1999年加入微软研究院历任数职，并在2014年初创办深度学习技术中心，主持微软公司和研究院的人工智能和深度学习领域的技术创新。邓立博士的研究方向包括自动语音与说话者识别、口语识别与理解、语音-语音翻译、机器翻译、语言模式、统计方法与机器学习、听觉和其他生物信息处理、深层结构学习、类脑机器智能、图像语言多模态深度学习，商业大数据深度分析等。他在上述领域做出了重大贡献，是ASA（美国声学学会）会士、IEEE（美国电气和电子工程师协会）会士和理事、ISCA（国际语音通信协会）会士，并凭借在深度学习与自动语音识别方向做出的杰出贡献荣获2015年度IEEE信号处理技术成就奖。同时，他也曾在顶级杂志和会议上发表过与上述领域相关的300余篇学术论文，出版过5部著作，发明及合作发明了超过70多项专利。邓立博士还担任过IEEE信号处理杂志和《音频、语音与语言处理学报》（IEEE/ACM Transactions on Audio, Speech & Language Processing）的主编。

俞凯

IEEE高级会员，上海交通大学计算机科学与工程系特别研究员。清华大学本科、硕士，英国剑桥大学工程系博士。长期从事智能语音及语言处理、人机交互、模式识别及机器学习的研究和产业化工作。他是中组部“千人计划”（青年项目）获得者，国家自然科学基金委优秀青年科学基金获得者，上海市“东方学者”特聘教授；作为共同创始人和首席科学家创立“苏州思必驰信息科技有限公司”。现任中国声学学会语音语言、听觉及音乐分会执委会委员，中国计算机学会人机交互专委会委员，中国语音产业联盟技术工作组副组长。他的研究兴趣涉及语音识别、语音合成、口语理解、对话系统、认知型人机交互等智能语音语言处理技术的多个核心技术领域，在本领域的一流国际期刊和会议上发表论文80余篇，申请专利10余项，取得了一系列研究、工程和产业化成果。在InterSpeech及IEEE Spoken Language Processing等国际会议上获得3篇国际会议优秀论文奖，获得国际语音通信联盟（ISCA）2013年颁发的2008-2012 Computer Speech and Language 最优论文奖。受邀担任InterSpeech 2009 语音识别领域主席、EUSIPCO 2011/EUSIPCO 2014 语音处理领域主席、InterSpeech 2014 口语对话系统领域主席等。他负责搭建或参与搭建的大规模连续语音识别系统，曾获得美国国家标准局（NIST）和美国国防部内部评测冠军；作为核心技术人员，负责设计并实现的认知型统计对话系统原型，在CMU组织的2010年对话系统国际挑战赛上获得了可控测试的冠军。作为项目负责人或Co-PI，他主持了欧盟第7框架PARLANCE、国家自然科学基金委、上海市教委、经信委，以及美国通用公司、苏州思必驰信息科技有限公司的一系列科研及产业化项目。2014年，因在智能语音技术产业化方面的贡献，获得中国人工智能学会颁发的“吴文俊人工智能科学技术奖”。

钱彦旻

上海交通大学计算机科学与工程系助理研究员，博士。分别在2007年6月和2013年1月于华中科技大学和清华大学获得工学学士和工学博士学位。2013年4月起，任上海交通大学计算机科学与工程系研究员。同时从2015年1月至2015年12月，在英国剑桥大学工程系机器智能实验室语音组进行访问，作为项目研究员与语音识别领域的著名科学家Phil Woodland教授和Mark Gales教授开展合作研究。现为IEEE、ISCA会员，同时也是国际开源项目Kaldi语音识别工具包开发的项目组创始成员之一。此外，担任IEEE Transactions on Audio, Speech, and Language Processing、SpeechCommunication、ICASSP、Interspeech、ASRU等国际期刊和会议的审稿人。目前在国内外学术刊物和会议上发表学术论文50余篇，Google Scholar总引用数近1000次。其中包括在语音识别领域权威国际会议ICASSP、InterSpeech

《解析深度学习：语音识别实践》

和ASRU上发表论文30余篇，申请国家专利共3项，已授权1项。2008年获科技奥运先进集体奖，2014年获中国人工智能学会颁发的"吴文俊人工智能科学技术奖进步奖"。曾作为负责人和主要参与者参加了包括英国EPSRC、国家自然科学基金、国家863等多个项目。目前的研究领域包括：语音识别、说话人和语种识别、自然语言理解、深度学习建模、多媒体信号处理等。

书籍目录

译者序	iv
序	vii
前言	ix
术语缩写	xxii
符号	xxvii
第 1 章 简介	1
1.1 自动语音识别：更好的沟通之桥	1
1.1.1 人类之间的交流	2
1.1.2 人机交流	2
1.2 语音识别系统的基本结构	4
1.3 全书结构	6
1.3.1 第一部分：传统声学模型	6
1.3.2 第二部分：深度神经网络	6
1.3.3 第三部分：语音识别中的 DNN-HMM 混合系统	7
1.3.4 第四部分：深度神经网络中的表征学习	7
1.3.5 第五部分：高级的深度模型	7
第一部分 传统声学模型	9
第 2 章 混合高斯模型	11
2.1 随机变量	11
2.2 高斯分布和混合高斯随机变量	12
2.3 参数估计	14
2.4 采用混合高斯分布对语音特征建模	16
第 3 章 隐马尔可夫模型及其变体	19
3.1 介绍	19
3.2 马尔可夫链	21
3.3 序列与模型	22
3.3.1 隐马尔可夫模型的性质	23
3.3.2 隐马尔可夫模型的仿真	24
3.3.3 隐马尔可夫模型似然度的计算	24
3.3.4 计算似然度的高效算法	26
3.3.5 前向与后向递归式的证明	27
3.4 期望最大化算法及其在学习 HMM 参数中的应用	28
3.4.1 期望最大化算法介绍	28
3.4.2 使用 EM 算法来学习 HMM 参数——Baum-Welch 算法	30
3.5 用于解码 HMM 状态序列的维特比算法	34
3.5.1 动态规划和维特比算法	34
3.5.2 用于解码 HMM 状态的动态规划算法	35
3.6 隐马尔可夫模型和生成语音识别模型的变体	37
3.6.1 用于语音识别的 GMM-HMM 模型	38
3.6.2 基于轨迹和隐藏动态模型的语音建模和识别	39
3.6.3 使用生成模型 HMM 及其变体解决语音识别问题	40
第二部分 深度神经网络	43
第 4 章 深度神经网络	45
4.1 深度神经网络框架	45
4.2 使用误差反向传播来进行参数训练	48
4.2.1 训练准则	48
4.2.2 训练算法	49

4.3 实际应用	53
4.3.1 数据预处理	54
4.3.2 模型初始化	55
4.3.3 权重衰减	55
4.3.4 丢弃法	56
4.3.5 批量块大小的选择	58
4.3.6 取样随机化	59
4.3.7 惯性系数	60
4.3.8 学习率和停止准则	61
4.3.9 网络结构	62
4.3.10 可复现性与可重启性	62
第5章 高级模型初始化技术	65
5.1 受限玻尔兹曼机	65
5.1.1 受限玻尔兹曼机的属性	67
5.1.2 受限玻尔兹曼机参数学习	70
5.2 深度置信网络预训练	73
5.3 降噪自动编码器预训练	76
5.4 鉴别性预训练	78
5.5 混合预训练	78
5.6 采用丢弃法的预训练	79
第三部分 语音识别中的深度神经网络 – 隐马尔可夫混合模型	81
第6章 深度神经网络 – 隐马尔可夫模型混合系统	83
6.1 DNN-HMM 混合系统	83
6.1.1 结构	83
6.1.2 用 CD-DNN-HMM 解码	85
6.1.3 CD-DNN-HMM 训练过程	86
6.1.4 上下文窗口的影响	88
6.2 CD-DNN-HMM 的关键模块及分析	90
6.2.1 进行比较和分析的数据集和实验	90
6.2.2 对单音素或者三音素的状态进行建模	92
6.2.3 越深越好	93
6.2.4 利用相邻的语音帧	94
6.2.5 预训练	95
6.2.6 训练数据的标注质量的影响	95
6.2.7 调整转移概率	96
6.3 基于 KL 距离的隐马尔可夫模型	96
第7章 训练和解码的加速	99
7.1 训练加速	99
7.1.1 使用多 GPU 流水线反向传播	100
7.1.2 异步随机梯度下降	103
7.1.3 增广拉格朗日算法及乘子方向交替算法	106
7.1.4 减小模型规模	107
7.1.5 其他方法	108
7.2 加速解码	109
7.2.1 并行计算	109
7.2.2 稀疏网络	111
7.2.3 低秩近似	113
7.2.4 用大尺寸 DNN 训练小尺寸 DNN	114
7.2.5 多帧 DNN	115

第 8 章 深度神经网络序列鉴别性训练	117
8.1 序列鉴别性训练准则	117
8.1.1 最大相互信息	118
8.1.2 增强型 MMI	119
8.1.3 最小音素错误/状态级最小贝叶斯风险	120
8.1.4 统一的公式	121
8.2 具体实现中的考量	122
8.2.1 词图产生	122
8.2.2 词图补偿	123
8.2.3 帧平滑	125
8.2.4 学习率调整	125
8.2.5 训练准则选择	126
8.2.6 其他考量	126
8.3 噪声对比估计	127
8.3.1 将概率密度估计问题转换为二分类设计问题	127
8.3.2 拓展到未归一化的模型	129
8.3.3 在深度学习网络训练中应用噪声对比估计算法	130
第四部分 深度神经网络中的特征表示学习	133
第 9 章 深度神经网络中的特征表示学习	135
9.1 特征和分类器的联合学习	135
9.2 特征层级	136
9.3 使用随意输入特征的灵活性	140
9.4 特征的鲁棒性	141
9.4.1 对说话人变化的鲁棒性	141
9.4.2 对环境变化的鲁棒性	142
9.5 对环境的鲁棒性	144
9.5.1 对噪声的鲁棒性	145
9.5.2 对语速变化的鲁棒性	147
9.6 缺乏严重信号失真情况下的推广能力	148
第 10 章 深度神经网络和混合高斯模型的融合	151
10.1 在 GMM-HMM 系统中使用由 DNN 衍生的特征	151
10.1.1 使用 Tandem 和瓶颈特征的 GMM-HMM 模型	151
10.1.2 DNN-HMM 混合系统与采用深度特征的 GMM-HMM 系统的比较	154
10.2 识别结果融合技术	156
10.2.1 识别错误票选降低技术 (ROVER)	157
10.2.2 分段条件随机场 (SCARF)	159
10.2.3 最小贝叶斯风险词图融合	160
10.3 帧级别的声学分数融合	160
10.4 多流语音识别	161
第 11 章 深度神经网络的自适应技术	165
11.1 深度神经网络中的自适应问题	165
11.2 线性变换	167
11.2.1 线性输入网络	167
11.2.2 线性输出网络	167
11.3 线性隐层网络	169
11.4 保守训练	170
11.4.1 L ₂ 正则项	171
11.4.2 KL 距离正则项	171
11.4.3 减少每个说话人的模型开销	173

11.5 子空间方法	175
11.5.1 通过主成分分析构建子空间	175
11.5.2 噪声感知、说话人感知及设备感知训练	176
11.5.3 张量	180
11.6 DNN 说话人自适应的效果	181
11.6.1 基于 KL 距离的正则化方法	181
11.6.2 说话人感知训练	183
第五部分 先进的深度学习模型	185
第 12 章 深度神经网络中的表征共享和迁移	187
12.1 多任务和迁移学习	187
12.1.1 多任务学习	187
12.1.2 迁移学习	189
12.2 多语言和跨语言语音识别	189
12.2.1 基于 Tandem 或瓶颈特征的跨语言语音识别	190
12.2.2 共享隐层的多语言深度神经网络	191
12.2.3 跨语言模型迁移	194
12.3 语音识别中深度神经网络的多目标学习	197
12.3.1 使用多任务学习的鲁棒语音识别	197
12.3.2 使用多任务学习改善音素识别	198
12.3.3 同时识别音素和字素 (graphemes)	199
12.4 使用视听信息的鲁棒语音识别	199
第 13 章 循环神经网络及相关模型	201
13.1 介绍	201
13.2 基本循环神经网络中的状态-空间公式	203
13.3 沿时反向传播学习算法	204
13.3.1 最小化目标函数	205
13.3.2 误差项的递归计算	205
13.3.3 循环神经网络权重的更新	206
13.4 一种用于学习循环神经网络的原始对偶技术	208
13.4.1 循环神经网络学习的难点	208
13.4.2 回声状态 (Echo-State) 性质及其充分条件	208
13.4.3 将循环神经网络的学习转化为带约束的优化问题	209
13.4.4 一种用于学习 RNN 的原始对偶方法	210
13.5 结合长短时记忆单元 (LSTM) 的循环神经网络	212
13.5.1 动机与应用	212
13.5.2 长短时记忆单元的神经元架构	213
13.5.3 LSTM-RNN 的训练	214
13.6 循环神经网络的对比分析	214
13.6.1 信息流方向的对比：自上而下还是自下而上	215
13.6.2 信息表征的对比：集中式还是分布式	217
13.6.3 解释能力的对比：隐含层推断还是端到端学习	218
13.6.4 参数化方式的对比：吝啬参数集合还是大规模参数矩阵	218
13.6.5 模型学习方法的对比：变分推理还是梯度下降	219
13.6.6 识别正确率的比较	220
13.7 讨论	221
第 14 章 计算型网络	223
14.1 计算型网络	223
14.2 前向计算	224
14.3 模型训练	227

14.4 典型的计算节点	231
14.4.1 无操作数的计算节点	232
14.4.2 含一个操作数的计算节点	232
14.4.3 含两个操作数的计算节点	237
14.4.4 用来计算统计量的计算节点类型	244
14.5 卷积神经网络	245
14.6 循环连接	248
14.6.1 只在循环中一个接一个地处理样本	249
14.6.2 同时处理多个句子	251
14.6.3 创建任意的循环神经网络	252
第 15 章 总结及未来研究方向	255
15.1 路线图	255
15.1.1 语音识别中的深度神经网络启蒙	255
15.1.2 深度神经网络训练和解码加速	258
15.1.3 序列鉴别性训练	258
15.1.4 特征处理	259
15.1.5 自适应	260
15.1.6 多任务和迁移学习	261
15.1.7 卷积神经网络	261
15.1.8 循环神经网络和长短时记忆神经网络	261
15.1.9 其他深度模型	262
15.2 技术前沿和未来方向	262
15.2.1 技术前沿简析	262
15.2.2 未来方向	263
参考文献	267

《解析深度学习：语音识别实践》

精彩短评

- 1、萝卜的老板
- 2、有人说阳春白雪，，真是侮辱了阳春白雪。也许作者真的很牛逼，但是牛逼的作者真不代表他的书是牛逼的。不建议阅读。其余的真的懒得说了。

《解析深度学习：语音识别实践》

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:www.tushu000.com