“十二五”国家重点图书

Springer 精选翻译图书

多模交互模式识别与应用

Multimodal Interactive Pattern Recognition and Applications

[西班牙] Alejandro Héctor Toselli
Enrique Vidal
Francisco Casacuberta 著

叶亮 马婷 译

哈尔滨工业大学出版社
HARBIN INSTITUTE OF TECHNOLOGY PRESS

“十二五”国家重点图书
Springer 精选翻译图书

多模交互模式识别与应用

Multimodal Interactive Pattern Recognition and Applications

[西班牙] Alejandro Héctor Toselli
Enrique Vidal 著
Francisco Casacuberta

叶亮 马婷 译

哈爾濱工業大學出版社
HARBIN INSTITUTE OF TECHNOLOGY PRESS

内 容 简 介

本书深入浅出地对多模交互模式识别技术进行了介绍，条理清晰，内容全面；对交互式模式识别的体系结构和性能特点进行了详细论述；通过多种应用系统实例对交互式模式识别技术进行了细致的分析，所涉及应用领域包括手写文档/语音转录、机器翻译、文本生成、图像检索等；而且书中的多数实例都在 Internet 上开放，读者可以下载系统原型，亲自动手实验，对交互式模式识别技术的原理和运用有更深入的了解。

本书始终以实例结合原理讲述交互式模式识别中的新算法，不论是作为研究生教学的教科书还是研发的工具书，都有很好的参考价值。

黑版贸审字 08－2016－114 号

图书在版编目(CIP)数据

多模交互模式识别与应用/叶亮，马婷译. —哈尔滨：哈尔滨工业大学出版社，2017.6
ISBN 978－7－5603－6191－8

Ⅰ.①多… Ⅱ.①叶… ②马… Ⅲ.①模式识别－研究 Ⅳ.①TP391.4

中国版本图书馆 CIP 数据核字(2016)第 220248 号

电子与通信工程
图书工作室

责任编辑 李长波
封面设计 高永利
出版发行 哈尔滨工业大学出版社
社　　址 哈尔滨市南岗区复华四道街 10 号　邮编 150006
传　　真 0451－86414749
网　　址 http://hitpress.hit.edu.cn
印　　刷 哈尔滨市石桥印务有限公司
开　　本 660mm×980mm　1/16　印张 18　字数 325 千字
版　　次 2017 年 6 月第 1 版　2017 年 6 月第 1 次印刷
书　　号 ISBN 978－7－5603－6191－8
定　　价 48.00 元

译者序

本书深入浅出地对多模交互模式识别技术进行了介绍，条理清晰，内容全面；对交互式模式识别的体系结构和性能特点进行了详细论述；通过多种应用系统实例对交互式模式识别技术进行了细致的分析，所涉及应用领域包括手写文档/语音转录、机器翻译、文本生成、图像检索等；而且书中的多数实例都在 Internet 上开放，读者可以下载系统原型，亲自动手实验，对交互式模式识别技术的原理和运用有更深入的了解。本书始终以实例结合原理讲述交互式模式识别中的新算法，不论是作为研究生教学的教科书还是研发的工具书，都有很好的参考价值。

本书的翻译由哈尔滨工业大学电子与信息工程学院的叶亮和哈尔滨工业大学(深圳)的马婷完成。本书共分为 12 章，其中第 1、2、5～9 章由叶亮翻译，第 3、4、10～12 章由马婷翻译。叶亮负责全书的统稿、修改与校对工作，并对原书中存在的疏漏进行了修订。本书的出版尤其要感谢李月、石硕、于启月、许震宇、朱师妲、高书莹、王鹏、任浩、任千尧、孙裕人、张佳伟、于婷、李亚添，他们在专业术语翻译、公式符号录入以及校对等方面花费了大量的时间和精力。没有他们的辛勤工作和严谨态度，就不能保证本书在这么短的时间内与广大读者见面。

本书的翻译是在国家自然科学基金(61602127)和教育部留学回国人员科研启动基金支持下完成的，特此感谢；还要感谢哈尔滨工业大学提供的各种设施，保证了本书翻译所需的各种资源。

由于译者水平有限，翻译过程中的疏漏和不当之处还请读者不吝指正，以便我们在下一版中进行改进。

译　者

2016 年 5 月 24 日

序

一般来说，模式识别的目的是自动解决复杂的识别问题。然而我们在很多实际应用中发现，全自动的识别系统很难达到所要求的正确识别率，因此经常需要用一些人工后期处理来修正识别系统产生的错误。但是另一方面，这些后期处理过程有时又会成为识别系统的瓶颈，因为它们会引入操作开销。

与其他模式识别方面的书籍相比，本书有两个特点。其一，本书使用一种完全不同的方法来修正识别系统产生的错误。该方法将用户和识别系统紧密地结合在一起，用户不仅要参与识别结果的修正，还要参与到识别过程之中。因此，很多错误就可以提前避免，从而节省了修正工作。其二，本书提出了多模人机交互的概念，用来修正和防止识别错误。这种多模交互除了传统的键鼠输入模式之外，还包括手写、语音、手势等输入模式。

本书中的素材都是基于成熟的数学理论，其中大部分是基于贝叶斯理论。书中包含很多多模交互模式识别这一新兴领域中的原创成果，并且针对一些具体应用展开了细致讨论。结果表明，相对于机器自动识别—用户后期处理的传统方法，多模交互识别系统具有很大的优势和发展潜力。本书的应用实例包括手写文本识别、语音识别、机器翻译、文本预测、图像检索和解析。

总之，本书给出了模式识别领域中一个非常新颖的观点，据我所知，这是第一本以统一、集成的方式介绍多模交互模式识别技术的书籍。这本书可能会成为这个新兴领域中的一个标准参考文献。我特别将这本书推荐给从事模式识别研究的研究生、学术或工程的研究者以及从业人员。

瑞士，伯恩
Horst Bunke

前 言

我们对人机交互的关注始于 TT2 项目("Trans-Type－2",2002～2005,*http://www.tt2.atosorigin.es*),该项目由欧盟(European Union,EU)资助,Atos Origin 协调,研究基于统计技术的计算机辅助翻译。

若干年前,我们完成了一个欧盟资助的语音机器翻译项目(EuTrans,1996～2000,*http://prhlt.iti.es/w/eutrans*),到 TT2 项目开始时,我们已经积累了几年的机器翻译(machine translation,MT)经验。因此我们很清楚机器翻译技术在实际应用时的关键瓶颈在哪里:当需要纠正翻译结果中的(大量)错误时,很多专业的翻译员宁愿自己从头输入所有的文本,也不愿意利用翻译系统正确翻译的(少量)单词。显然,在对 MT 系统给出的错误百出的翻译结果进行后期编辑时,这些专业人士并不认为是他们在掌控着翻译进程,反而更像是在给一个愚蠢的系统打下手——系统总是给出一些奇怪的翻译结果,而他们只能考虑如何去进行补救(这些年来后期编辑的方式可能得到了改进,但这种翻译过程完全"失控"的感觉仍然存在)。

在 TT2 项目研发过程中我们发现,如果能够充分考虑系统开发时所依照的数学公式,那么用户的反馈在计算机辅助技术的开发中可以起到核心作用,而且可以很大程度地提高系统性能。同时我们也注意到,传统的基于识别率的评价体系已不能全面地评价这些辅助技术,还需要有关人机交互工作量的评估指标。总之,计算机辅助技术的发展必须要遵循这样的原则:用户觉得系统的运行在其掌控之下,而不再是被动的修修补补,而且在设计系统时,也要把用户工作量指标作为一个基本出发点。在 TT2 项目中我们还意识到,多模处理实际上在所有的交互系统中都是存在的,并且可以利用多模处理来提升系统的整体性能和可用性。

继 TT2 成功之后,我们的研究团队(PRHLT—*http://prhlt.iti.upv.es*)开始研究,如何将这些理念应用到更多需要辅助技术的

模式识别领域之中。不久之后我们就又组织了一个庞大的西班牙研究工程，名为“模式识别与计算机视觉中的多模交互”(MIPRCV，2007～2012，*http://miprcv.iti.upv.es*)。这项工程的研究团队由来自 10 个科研院所的超过 100 名优秀博士构成，目的是开发交互式应用领域中的计算机辅助核心技术，例如语言和音乐处理、医学图像识别、生物识别和监控、先进驾驶辅助系统、机器人，等等。

本书的大部分内容是 MIPRCV 项目中 PRHLT 研究团队的研究成果，因此我们感谢所有直接或间接为本书贡献理念、探讨及技术合作的 MIPRCV 研究者，同时也感谢将其实现的所有 PRHLT 成员。

本书基于统计决策论的模式识别框架，以统一的形式介绍这些研究成果。首先，本书介绍了多模交互建模与搜索(或推导)的基本概念和通用方法；然后，给出一些基于这些概念和方法的具体应用系统，包括交互式手写文档或语音转录、计算机辅助翻译、交互式文本生成与解析、基于相关性的图像检索；最后，在最后一章给出这些应用系统的原型，其中大部分原型系统已有在线演示版本，并且在 Internet 上开放下载，读者可以亲自动手实验，学习如何在多模交互框架下应用模式识别技术。

本书内容如下：

第 1 章介绍了交互式模式识别的概念，讨论了在模式识别中引入用户交互框架的研究前景。本章基于现有的解决非交互式搜索问题的方法，给出了解决基本交互式搜索问题的通用方法，此外还给出了交互式框架下的现代机器学习方法总览。

第 2 章给出了计算机辅助转录技术(第 3～5 章内容)的通用基础和基本框架。第 3 章和第 5 章讨论了手写文档转录的不同方法，涉及方面包括多模态、用户交互方式和人体工程学、主动学习等。第 4 章则重点关注语音信号的转录，所用方法与第 3 章类似。

第 6 章介绍了交互式机器翻译，给出了一种人机交互框架，可以提高任意两种语言之间的翻译质量。这一章还展示了用户如何通过交互框架利用现有多模接口来提高系统效率。第 7 章和第 8 章分别讲解交互式机器翻译中的多模接口和自适应学习。

第 9～11 章讨论另外三种与前面技术大不相同的交互式模式识别课题：交互式解析、交互式文本生成和交互式图像检索。其中，交互式文本生成的特点是没有输入信号，而交互式解析和交互式图

像检索的特点是在分析输入信号时，使用的不是从左到右式协议。

最后，第 12 章给出了一些多模交互模式识别具体应用的系统原型演示，就如前面所说的，这些系统都是本书所介绍模式识别方法的应用实例，是真正可以用到实际系统中，实现人机交互的。

西班牙，瓦伦西亚
E. Vidal
A. H. Toselli
F. Casacuberta

术 语 表

缩写	全称	中文名称
3G	Third Generation (mobile networks)	第3代(移动网络)
3－gram		3词文法
AJAX	Asynchronous JavaScript and XML	异步Java脚本与XML
ANOVA	Analysis of Variance	方差分析
API	Application Programming Interface	应用程序接口
ASR	Automatic Speech Recognition	自动语音识别
ATROS	Automatically Trainable Recognizer of Speech	可自动训练的语音识别器
bi－gram		双词文法,或写作2－gram
CAT	Computer Assisted Translation	计算机辅助翻译
CATS	Computer Assisted Transcription of Speech	计算机辅助语音转录
CATTI	Computer Assisted Transcription of Text Images	计算机辅助文本图像转录
CBIR	Content Based Image Retrieval	基于内容的图像检索
CE	Confidence Estimation	置信估计
CER	Classification Error Rate	分类错误率
CKY	Cocke-Kasami-Younger(algorithm)	科克－卡西米－雅戈尔(算法)
CM	Confidence Measure	置信度
CNF	Chomsky Normal Form	乔姆斯基范式

缩写	全称	中文名称
CS	Cristo Salvador(corpus)	克里斯托萨尔瓦多(语料库)
CYK	Cooke-Yamakura-Kasami (algorithm)	库克－镰仓－卡萨米(算法)
DAG	Directed Acyclic Graph	有向无环图
DCT	Discrete Cosine Transform	离散余弦变换
DEC		(一种无约束语音解码方案)
DEC－PREF		(一种前缀约束语音解码方案)
DT	Decision Theory	决策论
ECP	Error Correcting Parsing	纠错解析
EFR	Estimated eFfort Reduction	预计工作减少量
EM	Expectation Maximization	期望最大化
ER	(classification) Error Rate	(分类)错误率
EU	European Union	欧盟
FER	Feedback decoding Error Rate	反馈解码错误率
FKI	Research Group on computer Vision and Artificial Intelligence	计算机视觉与人工智能研究小组
FS	Finite State	有限状态
GARF	Greedy Approximation Relevance Feedback	贪婪近似相关反馈
GARFs	Simplified GARF	贪婪近似相关反馈算法简化版
GIATI	Grammatical Inference Algorithms for Transducer Inference	用于转导推理的语法推理算法
GIDOC	Gimp－based Interactive transcription of old text DOCuments	基于 GIMP 的旧文本文档交互转录
GIMP	GNU Image Manipulation Program	GNU 图像处理程序

缩写	全称	中文名称
GPL	GNU General Public License	GNU 通用公共许可
GSF	Grammar Scale Factor	语法比例因子
HCI	Human-Computer Interaction	人机交互
HMM	Hidden Markov Model	隐马尔可夫模型
HTK	Hidden Markov model ToolKit	隐马尔可夫模型工具包
HTML	HyperText Markup Language	超文本标记语言
HTR	Handwritten Text Recognition	手写文本识别
HTTP	HyperText Transfer Protocol	超文本传输协议
IAM	Institute of Computer Science and Applied Mathematics	计算机科学和应用数学研究所
IAMDB	IAM Handwriting Database (hand-written English text)	IAM 手写数据库(手写英文文本)
IHT	Interactive Handwriting Transcription	交互式手写转录
IMT	Interactive Machine Translation	交互式机器翻译
IMT－PREF		(一种前缀约束语音解码方案)
IP	Interactive Parsing	交互式解析
IPP	Interactive Predictive Processing	交互式预测处理
IPR	Interactive Pattern Recognition	交互式模式识别
IRM	Independent Retrieval Method	独立检索方法
ITG	Interactive Text Generation	交互式文本生成
ITP	Interactive Text Prediction	交互式文本预测
JNI	Java Native Interface	Java 本地接口
KSR	Key Stroke Ratio	击键率
LOB	Lancaster-Oslo/Bergen	兰开斯特－奥斯陆/卑尔根
LSI	Latent Semantic Indexing	隐式语义索引
MERT	Minimum Error Rate Training	最小误差率训练

缩写	全称	中文名称
MFCC	Mel Frequency Cepstral Coefficients	梅尔频率倒谱系数
MI	Multimodal Interaction	多模交互
MIPR	Multimodal Interactive Pattern Recognition	多模交互模式识别
MM－CATTI	Multimodal Computer Assisted Transcription of Text Images	多模计算机辅助文本图像转录
MM－IHT	Multimodal Interactive Handwriting Transcription	多模交互手写转录
MM－IMT	Multimodal Interactive Machine Translation	多模交互机器翻译
MM－IP	Multimodal Interactive Parsing	多模交互解析
MM－IST	Multimodal Interactive Speech Transcription	多模交互语音转录
MM－ITG	Multimodal Interactive Text Generation	多模交互文本生成
MT	Machine Translation	机器翻译
n－gram		n 词文法
NLP	Natural Language Processing	自然语言处理
ODEC－M3	Spontaneous Handwritten Paragraphs Corpus	自发手写段语料库
OOV	Out of Vocabulary	超出词汇库
PA	Pointer Action	指针动作
PA－CATTI	Computer Assisted Transcription of Text Images using Pointer Actions	使用指针动作的计算机辅助文本图像转录
PAR	Pointer－Action Rate	指针动作率
PCFG	Probabilistic Context Free－Grammars	概率上下文无关文法

缩写	全称	中文名称
PDA	Personal Digital Assistant	个人数字助理
POI	Probability of Improvement	改进的可能性
POS	Part of Speech	词性
PR	Pattern Recognition	模式识别
PRCFG	Probabilistic Context Free－Grammars	概率上下文无关文法
PWECP	Probabilistic Word Error Correcting Parsing	概率字纠错解析
REA	Recursive Enumeration Algorithm	递归枚举算法
RF	Relevance Feedback	相关反馈
RISE	Relevant Image Search Engine	相关图片搜索引擎
rWER	residual WER	剩余 WER
SER	Sentence Error Rate	语句错误率
SFST	Stochastic Finite－State Transducers	随机有限状态转换器
SLM	Suffix Language Model	后缀语言模型
SMT	Statistical Machine Translation	统计机器翻译
SRI	Stanford Research Institute	斯坦福研究院
SRILM	SRI Language Modeling	SRI 语言模型
SVD	Singular Value Decomposition	奇异值分解
SVM	Support Vector Machine	支持向量机
TCAC	Tree Constituent Action Rate	树成分动作率
TCER	Tree Constituent Error Rate	树成分错误率
TCP	Transfer Control Protocol	传输控制协议
TMX	Translation Memory eXchange	翻译存储交换
TT2	TransType2	“TransType2”工程
UI	User-Interface	用户界面
uni－gram		单词文法，或写作 1－gram

缩写	全称	中文名称
WDAG	Weighted Directed Acyclic Graph	加权有向无环图
WER	Word Error Rate	误词率
WG	Word Graph	词图，或“字图”
WIP	Word Insertion Penalty	单词插入惩罚
WSJ	Wall Street Journal	Wall Street 期刊
WSR	Word Stroke Ratio	单词键入率

目 录