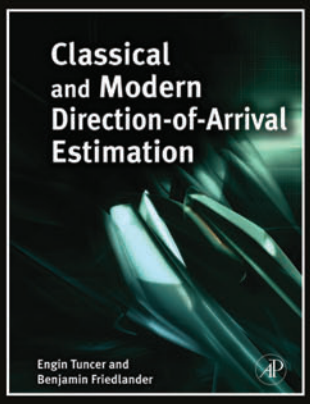
IEEE
Signal Processing
MAGAZINE

2010 年 IEEE 信号处理学会论文摘录

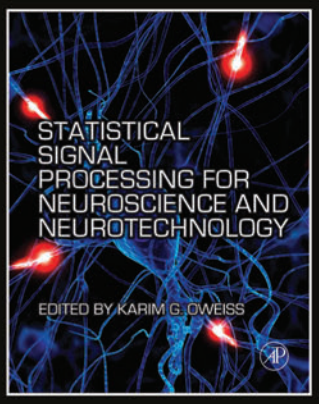
IEEE Signal Processing Society
IEEE

TASKED WITH SOLVING REAL-WORLD PROBLEMS?

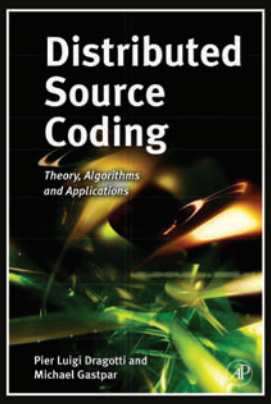
We have your answers – turn to Academic Press for advanced, up-to-date references covering all major areas of signal processing.



Classical and Modern Direction-of-Arrival Estimation
ISBN: 978-0-12-374524-8

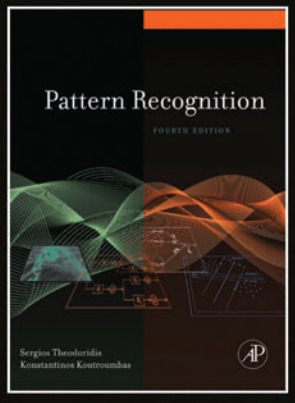


Statistical Signal Processing for Neuroscience and Neurotechnology
ISBN: 978-0-12-375027-3

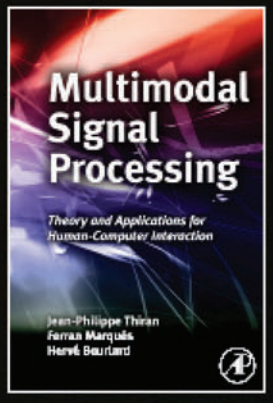


Distributed Source Coding
ISBN: 978-0-12-374485-2

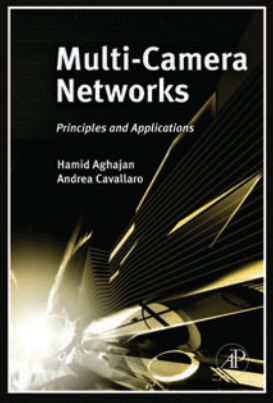
Handbook of Blind Source Separation
ISBN: 978-0-12-374726-6



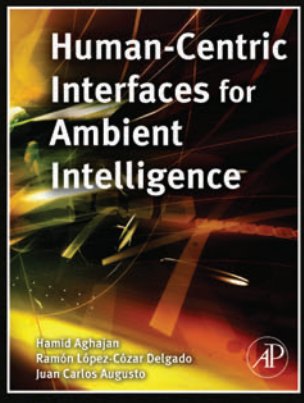
Pattern Recognition, 4th Edition
ISBN: 978-1-59749-272-0



Multimodal Signal Processing
ISBN: 978-0-12-374825-6



Multi-Camera Networks
ISBN: 978-0-12-374633-7



Human-Centric Interfaces for Ambient Intelligence
ISBN: 978-0-12-374708-2

All books are available as an e-book or print.

Find these and other great titles on elsevierdirect.com/engineering, Amazon.com or your favorite online retailer.

ACADEMIC PRESS

ELSEVIER

主编 邓立
deng@microsoft.com

拥抱信号处理的新黄金时代

信号处理对信息社会起到了至关重要的作用。信号处理随处可见：在手机、电视、汽车、GPS、调制解调器、扫描仪，以及各种各样的通讯系统和电子设备中都可看到信号处理的踪迹。现代手机的确是最有代表性的例子—语音、音频、图像、视频和图形都在这些小小的“奇迹”中进行处理和增强。这都有赖于数十年来人们对媒体信号处理的研究，而这些研究都曾在我们的 *IEEE 信号处理杂志（SPM）* 上出现。

近年来技术的进步已经预示着信号处理崭新的黄金时代的来临。有多个令人兴奋的方向，如生物信息学、人类语言技术、网络与安全，正在摆脱传统上对原始信息内容进行信号处理的领域。新时代的挑战则超越了处理低层次、波形类信号的传统角色，成为理解和挖掘高层次、以人为中心的语义信号与信息的新角色。在信号处理的某些领域已发生了这样的根本性转变。有望在未来数年的研究中，这样的转变在信号处理的更多研究领域中更加普遍。

除了传统的信号处理领域，如编码、分析、增强、合成、以及对普通的媒体和通讯信号识别外，*SPM* 还告诉我们的读者有关信号处理的新趋势。新技术的发展有赖于非传统的信号处理课题，其中包括对高层次信息源和内容的理解、挖掘，以及检索，这些往往都嵌入在低层次信号之中。为了让新的技术趋势形成更深刻的社会影响，在学术界与工业研究机构间，信号处理与其他相关学科间，我们需要比以往任何时候都强的互动。*SPM* 尤其是要促进这些互动。

在前任主编，张希福教授的领导下，凭借其编辑小组的勤奋工作，以及我们学会主管出版的副主席，刘国瑞教授所做的基础工作，*SPM* 状况颇佳，这有在全球 200 多个电气工程出版物中名列前茅为证。在新主编的过渡期，我得到了张教授和刘教授非常宝贵的指导和支持，对此我衷心感谢。同时我也很幸运，Antonio Ortega 教授、Dan Schonfeld 教授、Ghassan AlRegib 教授，以及吴旻教授分别同意担负起专题文章、专刊，专栏/论坛和电子通讯的领域编辑的责任。他们将与我和编辑部密切合作。我们将不负你们，也就是我们的读者的期望。

与其他出版物有所不同，我们的

数字对象标识符 10.1109/MSP.2008.930482

Digital Object Identifier 10.1109/MSP.2010.937894

杂志不聚焦于新的研究成果。我们的志专注于对重要理论、算法、工具、以及与信号处理相关的应用进行全面解读的指南性文章。IEEE 信号处理学会的每位会员每两个月就会收到一期 *SPM*。每期包括三个主要类别的文章—专刊文章、专题文章、和专栏/论坛文章。最近发行的信号处理内部电子通讯是SPM 每月推出的电子出版物，服务于学会的所有会员。

应对信号处理的新黄金时代的重大挑战并不容易。在将 *SPM* 打造为一个反映我们整个社团长期或短期的兴趣的杂志的过程中，我们寻求您的积极参与。如果您对于改善 *SPM* 中任何一个部分有任何想法，请联络我。我很欢迎您成为 *SPM* 的作者、专栏作家、客座编辑，和/或审阅人，并在信号与信息处理的崭新的，令人兴奋的时代中分享这个社团的乐观精神。

Ron Schneiderman

DSP 对新医学影像设计的影响

数字信号处理（DSP）对于医学影像现状的推进产生重大影响。

DSP 的优点众所周知：可以实时操作、高可靠性、非常节能。而且相对便宜。但是，医学影像市场还迫切需要更多的技术创新。更多关注的是更高的图像质量和设计更小型的系统。

“在未来数年，我们预期医学影像应用将实现从局限于基本的诊断功能传统影像准到一个小型化，高精度便携式医疗成像设备组成的新的生态系统的重大转变。”半导体市场研究公司 Databeans 的研究总监和总裁，Susie Inouye 说。

便携式系统的迅速发展促生了手持式设备，某些情况下，医疗和家庭监控设备更耐磨。在所有这些系统中，DSP 无处不在。因此，医疗设备制造商和芯片厂商都在努力扩大医疗诊断应用，并针对不断增长的市场引进新产品。

通用电气公司（GE）今年早些时候发布了其 vScan 扫描仪，大概有一个手机大小，售价不超过 10,000 美元。西门子已经升级了三年前首次推出的 Acuson P10 手持式扫描仪。

东芝公司最近也通过一个全新的笔记本电脑系统进入便携式超声波市场。日立公司也提供了一个笔记本电脑大小的系统。名为 Viamo，主要设计针对那些需要高端超声检查但却动弹不得的病人使用。

GE 传感与检测科技推出了一款轻量级（13磅）的便携式数字成像工具 DXR250V，其特点就是在这些先前仅限于计算机 X 线摄影和 X 线光片的应用中缩短扫描时间，最大程度地降低辐射照射。新的 GE 的设备可以连接到一台笔记本电脑产生用于即时检查图像。

一间很小的公司，Signostics 的一款手机大小，半磅重的 Signos 手持式超声波系统，近期获得美国食品和药物管理局（FDA）认证。

“Signostics 克服了开发手掌大小超声产品的设计上的艰难挑战”，Analog Devices 的医疗技术分部主任，Patrick O’Doherty说，该部门正在与 Signostics 紧密合作，为数据转换、信号调理和要实现其设计的传感器提供关键的信号处理技术。Signos 涵盖多个医学应用，包括腹部评估，如膀胱，腹主动脉瘤筛选和创伤评估，还有肌肉骨骼和基本产科。

SonoSite 公司作为床边检验市场的另一位参与者，提供了一个随身携带的超声波系统，主要是在医生的办公室中使用。

> 数字信号处理对于医学影像现状的推进产生重大影响。

Philips Medical 近10年前推出一款手持式超声设备。名为 OptiGo，该设备退出了市场，据说因为当时人们都质疑医疗成像设备的图像质量如此之差。

不同的系统、应用

目前有数个医学影像技术并存。

磁共振成像（MRI）提供了非常清晰的人体图像，用于疾病和伤害诊断的范围非常广泛。全世界每年进行超过 6000 万美元的 MRI 诊断程序。

MRI 是一种非侵入性技术，不使用磁共振电离辐射生成人体的图像。由于它能够定制检查，以满足诸如视场角的特定参数，对于很多不同的医疗状况，这是可选的诊断方法，包括癌肿瘤，韧带撕裂，和阿尔茨海默氏病。

电脑断层扫描（CT）是另一形式的扫描，产生人体内部器官的三维图像。随着技术的提高，CT 的使用越来越频繁，为内部器官、骨骼、软组织和血管的分析和诊断提供更清晰，更详细的图片“CT 扫描成像的进步从根本上改变了诊断成像的实际操作方法和经济学意义”，Databeans 的 Susie Inouye 说。（现在，在美国每年有超过 62 万名病人进行过 CT 扫描检查，而在 1980 年则是 3 万人。）

系统集成的进步有助于显着提升照片的数量（或“片数”），可采用 CT 机，提高图像的细节和质量。

数字化 X 射线是诊断技术从传统的 X 射线系统迈出的重要一步，其中每个组件的信号衰减消耗了原始 X 射线信号的 60%。通过增加一个用于数字 X 射线成像的数字探测器，可以捕获 80% 以上的原始图像信息。数字化的 X 射线辐射的使用也减少了病人的辐射剂量，通过去掉摄影处理过程，减少了诊断时间。高性能的数字信号处理器可以控制其功能和信号调节，以获取和改善数字 X 射线图像的清晰度。数字 X 射线的另一个重要优势是它能够存储和传输的数字图像。

诊断超声成像系统生成并传递声波，并捕获反射波，然后将其转换为可视化图象。对于接收到的声波所进行的信号处理过程包括插值、抽取、数据滤波和重建。可编程 DSP 和片上系统芯片（SoC）设计用来实现实时的复杂数学算法，以便有效地满足这些系统的处理需求。

另一种医学影像系统是正电子发射断层扫描（PET）。如同 MRI 一样，也是一种无创诊断技术。通过从身体（由病人食用的放射性化学元素所产生）辐射排放产生特定的器官或组织生理图像。

PET 系统通常使用 DSP 处理不同的输入放大器增益，并通过实际系统控制光电倍增管高压电源、探测环结

数字对象标识符 10.1109/MSP.2010.936828

合件的运动控制和病人出/入。DSP 也可以用于 PET 扫描仪控制和信号处理单元。

洛杉矶的 Westside Medical Associates 和比佛利山庄的 Westside Medical Imaging（WMI）最近报告说，早期的 PET 扫描可以在早期，可治疗阶段确诊老年痴呆症。“在纽约大学（NYU）Langone 医学中心的研究调查证实了我们长期持有的信念就是我们可以使用先进的成像技术在阿尔茨海默病人尚未出现症状的早期进行识别”，加州大学洛杉矶分校 Geffen 医学院的医学教授，WMI 的主任，Norman Lepor 博士说。

纽约大学的研究团队一直使用名为匹兹堡复合物 B 的荧光成像剂能够使乙型淀粉样蛋白斑块发光，这是发现阿兹海默氏症的一个特征。根据研究人员的研究，并不是大脑中存在乙型淀粉样蛋白斑的所有病人均发展为阿兹海默氏症。

西门子已经开发出一种新的成像系统，名为 Somatom Definition Flash 扫描仪，使用相对较低剂量的辐射，并只针对一个身体的特定区域进行扫描（参见“辐射照射可能要求设备的设计作出变更”）。

DSP 厂商利润的增长

几个主要的芯片公司正在努力推动改善医疗成像系统的准确度和效率的现状。

TI 长期以来一直是提供数字信号处理器和医疗成像应用相关设备的领军者，并于 2007 年成立了医学影像 DSP 部门。次年，又推出 1500 万美元的医学院校基金，希望在未来的三至五年内对医学技术产生的“重大影响”。

TI 的 DSP 医疗影像业务发展及市场部经理 Ken Nesteroff 说，超音波是 DSP 应用医学影像系统中较好的例证之一（参见图1）。

“当然，也有很多的模拟解决方案，我们会针对这些开发特定的部件”，Nesteroff 特别提到。“在 DSP 方面，我们把更多处理放在后端处理部分。您通常看到的是 B 超，彩超，多普勒功能中的千兆级 DSP，有时是射频信号解调。后端功能主要是将扫描转换为用于显示数据。在便携式系统中，该行业完全脱离 PC 机后端，而转向更多的片上系统方法”。

TI 目前正在改进其嵌入式处理器软件工具包，该工具包于 2009 年 3 月推出，用来协助医疗诊断超声厂商开发更精确和更经济的系统，而且速度很快。Nesteroff 说，新工具包的关键之处是在图像处理方面的进展。

TI 也看到了其最新的基于多核 DSP 的 SoC 架构在医学影像中的机会，该 DSP 合并有定点和浮点运算能力。专为通信基础设施设备所设计，新的 DSP 的运行速度高达 1.2 GHz，其引擎可提供高达每秒 2560 亿次乘法累加运算（256 GMACS）和每秒 1280 亿次浮点运算(128 GFLOPS)。

Analog Devices 作为医学影像行业的一个长期合作者，最近推出了一款新的电流/数字转换器芯片，使得高层数 CT 系统能够捕获实时的移动图像，如跳动的心脏，而且具有高精确度和细节信息。该芯片将光电二极管阵列信号转化为数字信号，根据 Analog Devices 的资料显示，与较旧的型号相比，可以将 CT 检测系统的电耗降低 50%，主要是通过更高度集成的设计来实现。

“需要铭记的是人设的成像系统都将作医疗诊断用，要保持图像质量，不能损失信息，医生才能进行辨别”，Analog Devices 医疗保健技术部门的战略营销经理 Tony Zarola 如是

辐射照射可能要求设备的设计作出变更

最近数月以来，对于病人而言，辐射风险已成为一个大问题，同时也成了各医疗成像系统制造商、放射科医师和医生的热门话题。

联邦监管机构认为必要的 CT 扫描可发现众多的健康问题，但他们同时也发现越来越多的证据表明暴露于辐射的人可能会增加其在未来罹患癌症的风险。

正因如此，美国 FDA 的医疗器械中心（CDRH）已经启动了一个倡议，这可能迫使成像设备制造商重新设计他们的产品，在辐射剂量超过建议的水平时，就可以提醒医护人员当。

FDA 在 4 月初举行的首个系列会议中讨论了如何保护病人免受不必要的辐射照射。

FDA 表示，其目标是支持医学成像所带来的好处，同时尽量降低风险。CDRH 的主任，Jeffrey Shuren 博士说：“美国人受到来自医疗成像的辐射照射，在过去 20 年大幅度增加”。事实上，最近的研究表明，过去30年来，普通美国人的总辐射照射量增长了近一倍，主要来自于 CT 扫描和其他下一代成像测试。

例如，腹部 CT 扫描的辐射剂量大约是胸部X光检查的 400 倍。相比之下，牙科 X 光检查需要大约胸部 X 光检查的辐射剂量的一半。FDA 声称，计划向 CT 和 X 光透视设备制造商发出针对性的需求，从而在他们的机器设备设计中纳入直观重要的防护措施，以便开发更安全的技术，并提供适当的培训，支持从业人员的安全使用。该机构于 3 月下旬举行首个系列公开听证会以听取需要设立什么样的要求。

为了赋予病人权利，并提高认识，FDA 正在与其他组织合作开发和传播患者医学影像历史记录卡。该工具可以从 FDA 的网站上获取，让病人追踪他们自己的医学影像历史，并与他们的医生分享，特别是当这些影像可能没有包含在他们的医疗记录内时。

医学影像与技术联盟（MITA），一个代表了医学影像和放射治疗系统制造商的协会，声称他们支持措施减少不必要的辐射照射，并最大程度减少医疗过失的倡议。

美国放射技师协会声称支持 MITA 在所有 CT 新产品中纳入辐射剂量检查功能方面所做的努力，儿科成像辐射安全联盟也一样，该联盟领导“温柔影像”运动，以减少接受医学影像检查的儿童所承受的辐射剂量。

说。更高的图像分辨率会转换成更多的像素，Zarola 说，这就意味着在更多的数据和更高后端图像处理要求。

一个显而易见的目标就是在扫描期间，获得更多图像信息的同时更少地照射（减少扫描时间）有害的 X 射线。就系统中的电子器件而言，更多扫描线就意味着需要更多通道，更高的图像分辨率就转化为更多的像素，而更高的信噪比的无线设备提供给更低的噪声，因而可获得更好的对比度。

“要传输更多来自接收器的数据，增加的通道数就需要增加整个系统的带宽。”，Zarola 补充道。“这可能会导致我们要面对在现有带宽有限的基础设施传输数据的挑战。”

DSP 的优势意义重大，他强调，范围从降低带宽到智能压缩算法的使用。（可以使用有损压缩，但随后对图像的完整性所造成的影响将需要特征化。）“为了获得更好的图像质量，可使用各种后处理图像增强算法能够提高对比度或减少系统噪声的影响”，Zaroloa 如是说。“同样，面临的挑战将是保持图像的完整性。”

一个巨大的市场

医学成像已经是一个巨大的市场，并持续不断地增长，在很大程度上得益于由于技术的进步，以及日益普及的便携式成像产品。据 Reportlinker 所作的市场研究表明，医学成像设备的全球市场预计到 2015 年将达到约 370 亿美元。

MRI 预计在 2005-2015 年期间增长速度最快的成像方式，其年复合增长率（CAGR）为9.8%。

另一家市场研究集团，Global Industry Analysts 声称，美国、日本和欧洲占有 CT 扫描仪全球市场用户基数的 85% 以上。据 Global Industry Analysts 的说法，全球 CT 扫描仪市场主要有四家公司：GE 医疗、西门子医疗、东芝医疗系统，以及飞利浦医疗保健。其他主要参与者包括日立医疗公司和日本岛津医疗系统。

技术的快速升级也对 CT 扫描仪产生了影响。在 CT 领域的主要趋势是朝着组合扫描仪的方向转变，主要是组合了 PET 和 CT 成像能力的复合扫描仪。

Global Industry Analysts 声称，自 50 年代初引进后，超声波赢得医疗影像市场越来越多的份额。超声设备小型化，并持续不断地将电子系统纳入超声技术是主流趋势，该成像技术的成功多归功与此。

超声设备的整体市场在美国已接近饱和；然而，心脏科还是继续体现了超声终端产品的快速增长，预计 2010 年在美国境内的收益将达到 6.84 亿美元。这个市场基本上是使用新的、技术上更先进、系统话的新产品更换和升级老化设备的需求所驱动。美国和欧洲加在一起，约占全球 60% 的医学超声设备市场份额，尽管根据市场调研公司的说法，亚太市场正在迅速崛起。

根据 Global Industry Analysts 的说法，医疗超声市场的主要参与者有 Aloka 公司、B-K Medical、Esaote SPA、GE 医疗集团、日立医疗系统美国公司、Medison 有限公司、飞利浦医疗、西门子医疗、SonoSite 公司、TomTec 成像系统公司、GmbH 和东芝医疗系统有限公司。据报道，飞利浦、西门子、通用电气和东芝占全球市场份额的 80%。

> 数字化的 X 射线辐射的使用也减少了病人的辐射剂量，通过去掉摄影处理过程，减少了诊断时间。

MRI 市场预计在 2010 年将达到的 5.5 亿美元，推出的高场系统和诸如功能性神经成像、磁共振血管造影术、无创结肠镜检查以及乳腺 MR 等新技术推动了这个市场。

MRI 设备的关键卖点似乎是较高图像质量和成本效益。根据 Global Industry Analysts 的说法，GE 医疗，西门子医疗系统，飞利浦医疗系统称霸全球 MRI 设备市场，其他知名公司包括 Esaote 公司、日立、东芝医疗系统公司、Fonar 公司、IMRIS 与 Medtronic。

研究了医学影像市场的另一个研究机构，Frost & Sullivan 最近发布了一份报告，提到在欧洲刮起了一阵有关医学成像的研究与发展（R&D）的风潮，特别是心脏科的应用。F&S 预期制造商在超声心动图方面还有巨大的市场机会，藉此可以为私人执业医生提供便携式、基于 PC 的超声系统。

巨大的 DSP 需求

数字信号处理进入医疗系统市场的切入点在那里？

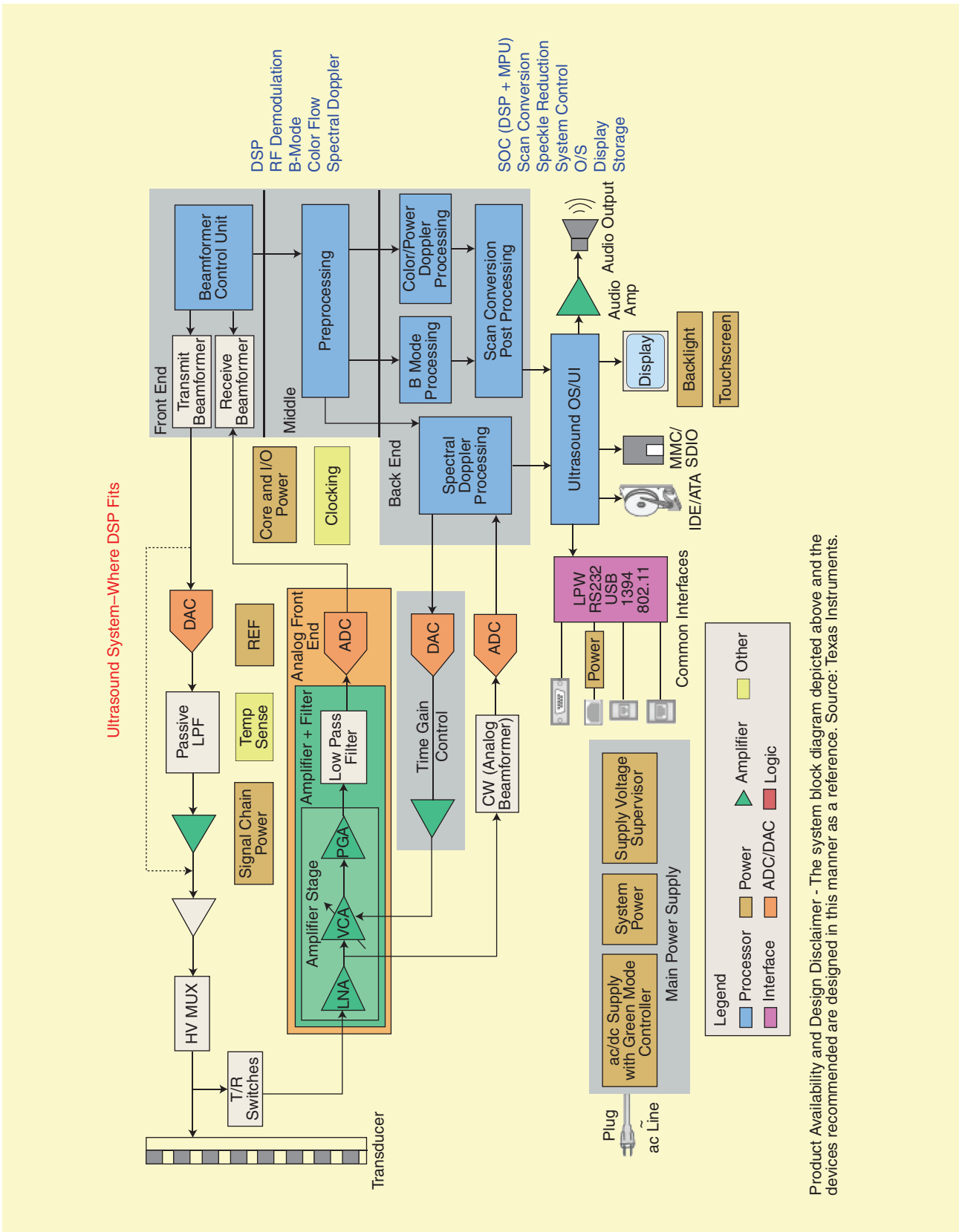
Databeans 预测，销往全球医疗成像应用的 DSP 的收益将增长近一倍，从 2008年 3140万美元增长到 2014 年的 6060 万美元（参见表1），主要是这些系统的市场有所增长，而且在数个方面的技术也都有所进步。

其中一个技术进步就是将 X 光片转化成了数字文件。DSP 有助于将捕获点的 X 射线信号转换成数字图像，而无需对图像完整性进行取舍。正如 TI 在一份有关医学成像未来的报告中所指出的那样，能够提供实时数字图像的能力使得数字 X 射线机可应用于外科手术，能够让医生在手术中看到精确的图像。

就在几年前，MRI 也进行了改进，能够在短的时间内提供更高质量的图像。此外，扩散 MRI 允许研究人员创建脑图谱，以便通过跟踪技术研究完全不同的大脑区域之间的关系。功能性 MRI 现在可以迅速扫描大脑以便测量因神经活动而引起的信号变化。DSP 也在远程医疗方面发挥着至关重要的作用，特别是视频会议和远程呈现系统中，可支持多种格式。

根据 TI 的报告中的陈述，DSP的使用是“贯穿这些实例的一个共同的主题”。更重要的是，这项技术对全球的医疗水平产生了意义深远的影响。

[SP]

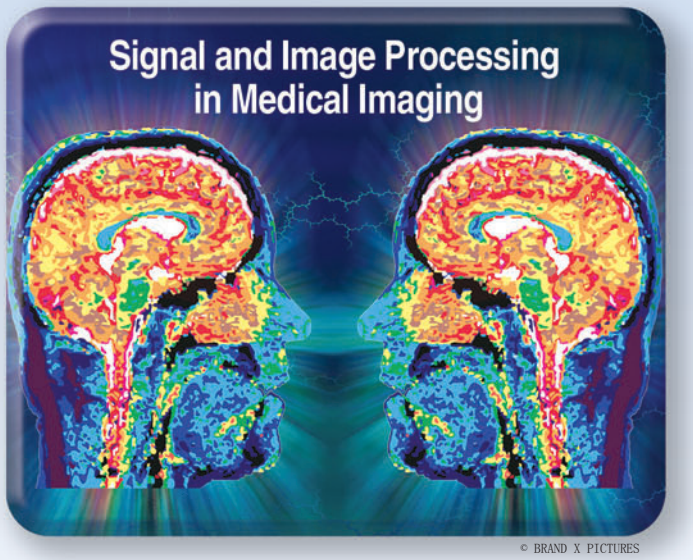


[图1] 该系统框图是 TI 针对 DSP 和 其他设备所建议的参考设计，可以用在超声医学成像系统的设计中。（图得到了使用许可。）

Miles N. Wernick Yongyi Yang Jovan G. Brankov
Grigori Yourganov Stephen C. Strother

医学成像中的机器学习

来自医学成像的结论



多个世纪以来，在许多领域，创造（和再创造）了多个自动决策和建模的统计方法。在这方面，重要的问题包括模式分类、回归、控制、系统辨识和预测。近年来，这些思想被确认为机器学习这个统一概念下的不同实例。这是有关 1）在现有数据内部量化关系的算法的发展，以及2）利用这些已识别的模式，并基于新数据作出预测。光学字符识别，即根据前面的实例自动识别印刷字符，就是一个机器学习的经典工程实例。但本文将讨论使用机器学习的非常不同的方式，您可能不是那么熟悉，我们将通过实例来演示这些概念在医学成像中所扮演的角色。

在现代计算的背景下，机器学习已引起了空前浓厚的兴趣，例如商业智能、检测垃圾邮件，以及欺诈和信用评分。相对于其他领域，在医疗成像领域采用现代机器学习技术则进展较慢。然而，由于计算机的能力越来越强大，因此对采用先进的算法颇有兴趣，可方便我们使用医学影像，并增强我们从中获取的信息。

尽管机器学习这个术语相对较新，机器学习的理念已应用到医学成像数十年，或许最引人注目的是在计算机辅助诊断（CAD）和脑功能活动定位领域。我们并不会在这篇简短的文章中回顾这个领域的丰富文献。我们的目标反而是1）介绍目前机器学习领域的一些主要的先进技术以飨读者，2）解释说明如何在医学成像中以不同的方式使用这些技术，其中使用了如下来自我们的研究的例子：

- CAD
- 基于内容的图像检索（CBIR）
- 图像质量自动评估
- 脑成像

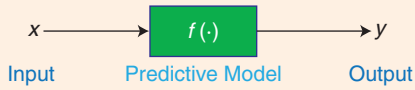
数字对象标识符 10.1109/MSP.2010.936730

机器学习简介

在这篇简短的指导性文章中，利用我们过去在这一领域的工作实例，尝试介绍一些广泛适用的基本技术，并说明这些技术被如何用于各种医学成像背景。如需更进一步的信息，有兴趣的读者可以参考那些有关机器学习的众所周知的介绍，如参考文献[1]和[2]中所作的精彩论述。

有监督学习

在机器学习中，人们往往旨在根据输入变量 x 预测输出变量 y。要达到这个目的，则假定输入和输出间基本服从一个函数关系，$y=f(x)$，称之为预测模型，如图1所示。在有监督学习中，借助于由 x 和 y 均已知的例子组成的训练数据发现预测模型。我们把所有可用的例子对记为 $(x_i\ y_i), i=1,...,N$。我们假设 x 由 个变量（称为特征）组成，那么 $x \in \mathbb{R}^n$ 。一般而言，该预测模型的输出可以是一个向量（例如在多类分类器中），但为了简单起见，我们将把注意力集中到标量输出的情况。

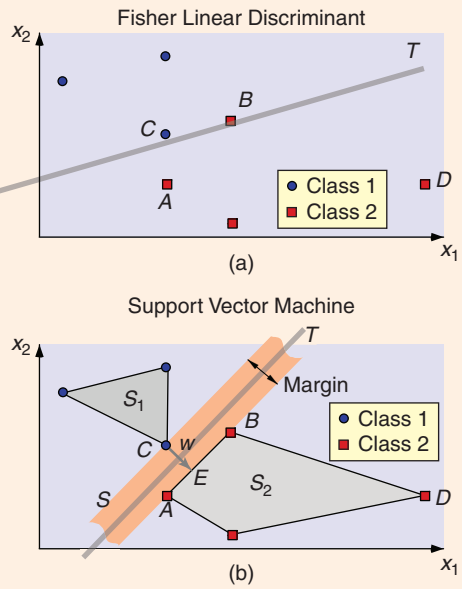


[图1] 在有监督学习中，预测模型表示输入变量x与输出变量y之间的假设关系。

从历史上来看，某种程度上来说，有时人为地将其分成两个学习问题：分类和回归。分类是指在特别小而离散的选择集合中做出决策（如识别肿瘤是恶性或良性），而回归则是对可能的连续值输出变量进行估计（如疾病严重程度的诊断评估Y）。如果一个分类问题中的选择由离散的数值表示（例如，$y=+1$代表恶性，$y=-1$则代表良性），那么就很容易地看到，图1中的模型则等同地表示了分类和回归。

支持向量机分类：最大间距法

让我们来看看图2中所描述的简单模式分类问题，其目的是使用决策边界T将向量$x=(x_1,x_2)^T$分成两类，我们使用线性模型$f(x)=w^Tx+b$，因此在这个两维的例子中，T是一条直线。传统上来说，模型的参数（此例中为w和b）通过诸如最小二乘或最大似然等经典准则确定。图2解释说明为什么这种做法很容易失败（在这种情况下，就是 Fisher 判别），特别是不符合该方法的分布假设时。在图2（a）中，数据点 D 对 Fisher 判别边界有不利的影响，而导致 B 点误判，即使是D点位于类别1很远的地方，不应受到如此程度的影响。



[图2] Fisher线性判别法（LD）和支持向量机。在这个例子中，（a）因为训练实例D对决策边界的不良影响，Fisher LD未能将其分成两类（b）SVM 则只使用点A、B和C，即所谓的支持向量，来定义决策边界，完全不受D点的影响。

而 Vapnik 提出支持向量机（SVM）[2]则解决了这一缺点，通过只考虑那些位于他们很接近不属于的类别的训练样本来确定判定边界。这一思想在如图2所示的情况下最容易理解，如图2所示，这两个类别通过一个线性决策边界严格可分，正如 Wernick 在文献[3]中所探讨的一样。在这种情况下，可将两个类别的间距最大化的分离线通常通过如下的方法寻找：

1) 画出数据点的类别凸包（就像环绕着每个数据点的群组撑开一个橡皮筋；称其为区域S1和S2）。

2) 寻找点 C 和 E，让区域 S1 和 S2 与他们之间的距离最小。

3) 画出连接点 C 和 E 的线段的垂直平分线，则就得到了决策边界。

第2步是通过使用采用标准方法[3]求解一个二次规划（约束优化）问题来完成的。在线性分类器中，向量w被称为判别向量。

在 SVM 的术语中，图2中的点A、B和C被称之为支持向量，这时一个类比于力学而衍生出来的术语。如果图2中的点A、B和C是物理支持，则足够为他们之间的夹心板提供足够的力学稳定性。

很显然，支持向量是来自训练数据，可以明确定义该模型的那些最适合的实例。确切地说，对于特定的测试例子，可以根据支持向量写出模型，如下所示：

$$f(x)=\sum_{i\in I_s}\alpha_i y_i x_i^T x+b \qquad (1)$$

其中仅将是支持向量的训练实例进行求和，并使用优化过程中的拉格朗日乘子确定系数。

支持向量机方法的好处在于分类学自动专注于难以归类的实例点A，B和C；式（1）中的计算结果随着支持向量的数量而改变，并不是因为空间维数的增加而变化（在某些问题中，空间维度非常大）。此外，可以证明支持向量机能够在训练误差和模型复杂性间取得平衡，从而避免过拟合，其中一个缺陷就是根据训练实例将模型微调得非常好，但对于新数据却无能为力。这种方法被称为结构风险最小化。

所描述的公式远远没有顾及到通过线性边界不能将两个类别完全分开的可能性。但是这种情况通过将松弛变量引入二次优化问题就很容易地得到解决，从而使错分的训练数据的数目降到最低。此外，支持向量机可以很容易地实现回归，而不是通过所谓的不敏感成本函数进行分类[2]。

非线性模型：核方法

机器学习的一个重要突破就是已获承认的所谓的核方法[2]，它提供了一个简单和广泛适用的手段，通过内积，从任意的线性模型获取一个非线性模型。即便是经典的方法，如 Fisher 判别或主成分分析，都可以通过核方法很容易地转化为灵活的非线性方法。

为了理解核方法，考虑应用以下的假设的一系列的步骤将线性 SVM 转化成为一个非线性技术。假设我们先对每一个来自训练集的输入向量x_i进行非线性变换Φ，并训练一个线性分类器用于区分变换后的向量$\Phi(x_i)$的类别。如果转换的空间维度较原来的空间高，可分性将会增强，转换维度的确不是有限的。

乍看之下，将每个输入向量转换到一个高维空间似乎是不切实际的。然而，核方法让我们认识到无需真正地进行变换就可用得到期望的结果。使用变换并应用于式（1）中的SVM 模型，就可以得到。变换后，（1）就变为：

$$f(x)=\sum_{i=1}^{N} w_i K(x,x_i) \tag{2}$$

请注意，（2）式中的变换Φ仅以内积$K(x_i,x)\triangleq\Phi(x_i)^T\Phi(x)$的形式出现，那么式（2）就可以写为：

$$f(x)=\sum_{i=1}^{N} w_i K(x,x_i) \tag{3}$$

因此，可以看到我们实际上没有必要去计算Φ（甚至不用明确定义）。相反，只需简单地定义 核函数$K(\cdot,\cdot)$ 而已，结果表明，任何半正定函数就足够了。机器学习中常用的核函数包括径向基函数（高斯）和多项式。直觉上来说，核的作用就是衡量测试向量x和每个支持向量x_i间的 相似性 ；这些相似性将用于获得输出结果。属于这些类别之一的向量可能是与属于这一类的支持向量最 相似 ，因此这些相似度就传达了所需要的信息。要记住的关键点是这些相似性比较只是相对于支持向量，对于靠近判定边界的实例就问题多多。稍后我们将看到在乳腺X光摄影背景下这些支持向量的视觉实例。

关联向量机：贝叶斯学习和稀疏约束

支持向量机的一个重要的改进就是 Tipping 提出的所谓的关联向量机（RVM）[5]。我们发现 RVM 在数个医学成像应用中的性能异乎寻常地良好，通常比其他可选择方法，包括 SVM，计算成本低很多。RVM 强调稀疏性（即降低模型的复杂度），因此与压缩感知的思想密切相关[6]。就像 SVM 一样，RVM 使用了被称为关联向量的训称数据子集，但关联向量通常要比支持向量少得多。

正如 SVM 一样，关联向量机也以核模型开始

$$f(x)=\sum_{i=1}^{N} w_i K(x,x_i) \tag{4}$$

然而，SVM 是以最大分类间隔准则为基础，而关联向量机则使用了贝叶斯方法。RVM 假设核加权w_i服从高斯先验分布，具有零均值和方差a_i^{-1}。RVM 则进一步假设aa_i^{-1}服从超 Gamma 先验分布。选择这些模型的直接后果就是核加权w_i总的先验分布为一个多元 t 分布。由于该分配是关于w_i空间的轴线紧密集的，那么先验分布的大部分的值都几乎为零。因此，最终求和只涉及到少数几个非零项w_i，相关的训练实例被称为相关向量。使用这个机制，通常可以避免过拟合，而且 RVM 的计算时间则相对较短。令人惊讶的，尽管具有这样的优势，RVM 在医学成像中的使用还相对较少，特别是与广为人知的 SVM 方法相比较。

虽然 RVM 和 SVM 都完全以训练数据的子集为基础给出决策（RVM 中使用相关向量，SVM 中使用支持向量），这些子集通常是完全不同的。支持向量都是位于决策边界的实例，而相关向量则通常传播到整个分布。稍后我们将在有关乳腺 X 光摄影的上下文中看到这种差异。

不幸的是，RVM 不像 SVM 那样有一个简单的几何解释，因此在本文中将不会给出一个图形方面的实例；我们请读者参考文献[5]，其中包含数个非常精彩的插图。

统计重采样的鲁棒性和评价

统计数据重采样[7]指的是用来评价机器学习模型的性能和改善鲁棒性，并估计统计显着性水平的系列技术。尽管重采样没有预测模型那么受关注，但也同样重要。

机器学习与经典决策和估计理论的主要不同之处在于它着重于在只有来自数据本身的数据基础分布的知识的状况。在这种环境下，统计显著性检验就不能使用传统的方式处理，因为零分布是未知的。幸运的是，零分布的经验估计可以很容易地通过置换重采样获取。

为了理解置换重采样，我们考虑如下状况，这里有两套数据，w_1和w_2，我们希望测试某些假设，例如他们的均值是否一致。由于我们不知道事实上w_1和w_2是否服从相同的分布（乃至分布形式），我们就无法直接评估显著性。但是我们可以通过重排数据上的标签，创建一个经验零分布，例如故意创建两个数据集，其中的数据来自w_1和w_2的混合组。请注意，往往重要的只是标签，而不是数据本身被重排（例如，在时间序列问题中）。通过以各个可能的方式（或至少以某些相当大量的随机方式）置换数据，我们就可以获取实例数据，其中我们知道这两个群体服从相同的分布，从而刻画了零假设。

重采样可发挥核心作用的另一个地方是在解决以下模型验证问题中：如果我们用所有可用的数据训练我们的模型，那么就没有留下测试模型或优化参数所需的数据。在这里使用置换重采样方法有交叉验证和 bootstrap 方法，都需要独立同分布（IID）的重采样对象。在k-折 交叉验证中，数据集被随机地分为k个组；其中的（k-1）个组用于训练模型，保留下来的那一组则用于测试。此过程需要进行k次（每个被拿出来的组都要执行一次），然后把结果组合起来，通常是求平均。在基本 bootstrap 法中，数据而是是用包含 N 个数据实例的集合进行训练，这些数据实例通过在整个数据集中随机置换 N 次的重采样获得。偶然有一些没有被选入训练集实例，则都将留给测试用。就像在交叉验证中一样，该过程将重复进行，结果则通过求平均进行结合。

对于基本 bootstrap 法而言，众所周知是以向下偏误为代价，可以降低估计方差的预测准确度（即基本 bootstrap 法给出了悲观的性能估计）。0.632 bootstrap法采用了偏差修正项对此进行了改进，在更先进的bootstrap 中[8]，尝试解释说明因过拟合产生的偏差。在使用置换获得经验零分布的问题中，替代假设的经验零分布通过可以通过 bootstrap 获得。

统计重采样不仅被广泛地用于测试预测模型，还可以改善其性能。有关的例子包括 bootstrap aggregation （即 bagging）技术，以及非参数，预测，活化，影响力，复现性，影像学中的重采样框架（NPAIRS）[9]，本文将在稍后作出解释。

乳腺 X 线影像诊断中的计算机辅助检测

CAD（computer aided detection，计算机辅助检测）在过去数十年都是非常活跃的研究领域，所以我们不打算在此对有关文献进行全面考察。感兴趣的读者可以参考有关计算机辅助检测乳腺X线影像的总结与回顾，比如文献[10]和[11]。

> 在现代计算的背景下，机器学习已引起了空前浓厚的兴趣，例如商业智能、检测垃圾邮件，以及欺诈和信用评分。

或许 CAD 最大的成功就在于乳腺成像。研究表明，如果有两个放射科医师判读同一个乳房X光检查结果，可明显提高癌症筛查的敏感度，但这将以增加工作量和成本为代价。CAD 软件可以用作一个替代的 第二读者 ，以更低的成本提高放射科医师的诊断准确性为目的。

计算机辅助检测的 CAD，即 CADe, 计算机会提醒放射科医师潜在的病灶：计算机辅助诊断，即 CADx，可预测病灶是恶性的可能性有多大。

CAD 体系通常包括下列主要步骤：1）应用自动图像分析，提取定量特征向量以便表征相关图片的内容，以及2）应用模式分类器，以确定所提取特征向量可能属于那个类别。

自动提取图像特征包括图像对比度，与几何形状、形态学和纹理特征。此外，还有针对该病人提供的其他形式的可用信息。机器学习的应用范围从线性判别（LD）分析、模糊逻辑技术、神经网络和委员会的机器，直到本文在前面所说明的最近的基于核的方法（如 SVM 和 RVM）。

接下来，我们将描述两个机器学习用于数字化乳腺 X 光筛查的例子，这两个例子均来自我们自己的研究工作：微钙化点簇的检测（CADe）和分类（CADx）。

CADe：微钙化检测

Microcalcifications （MCs）指乳腺组织内微小的钙沉积，在乳腺 x 光片上显示为小亮点（参见图3）。微钙化点簇是乳腺癌的重要指标，在30-50％的病例中出现。单个的 MCs 有时很难发现，因为其形状，方向，亮度和大小（通常为0.05-1mm）的变化，而且还因为周围的乳腺组织存在混杂纹理。微钙化检测一向是深入调查的目标（例如文献[12]）。已经证明现代机器学习方法对此非常有效，就如我们接下来要解释的那样。

SVM 检测器

在文献[13]中，我们训练了一个 SVM，依据围绕该点的较小感兴趣区域（ROI）为基础，用于确定乳腺 X 光影像的每个位置是否存在 MC（ MC 存在 类）或不存在（ MC 不存在 类）。SVM 通过放射专家所识别出来的 MC 存在 的 ROI 进行训练（参见图4）。

MC 通常只占据了一个乳腺X光片的一小部分，因此 MC 不存在 的 ROI 要比 MC 存在 的 ROI 多。要充分利用这一优势，我们开发了连续增强学习（SEL）中的流程，提高了 SVM 分类器的预测能力。在 SEL 中，通过从整个可用训练图像中选择最具代表性的 MC 不存在 实例，重迭调整 SVM 训练，同时保持训练实例的总数较小。

以受测乳腺 X 光影像集合为基础，通过衡量自由响应的受试者操作特性（FROC）曲线，我们可用说明了 SEL-SVM 方法在有文献可考的数个先进的方法中，性能最佳，检测概率与每幅图像假阳性的平均数量绘制了图5。图3给出了一个实例图像的部分，以及相应的 SVM 输出。

RVM 检测器

在乳腺 X 光筛查中，计算时间可以说是一个非常严峻的问题，其中的图像可以包含非常多的必须评估的 3000 5000 像素点。尽管 SVM 获得了非常卓越的检测性能，但是非常耗时，因为支持向量的数量可能十分巨大。为了解决这个问题，在文献[14]中，我们开发了一种基于 RVM（如前所述）的方法，该方法能够产生一个非常稀疏决策函数，从而显著节省计算时间，而同时也可以得到类似 SVM 的检测性能。

为了进一步加快算法，我们探索一个两阶段的分类方法，我们使用一个计算量较小的线性 RVM 作为第一阶段，用于快速消除非 MC 的像素，然后再使用一个非线性的 RVM 分类器检测中剩余的 MC。我们的研究结果表明，RVM 方法实现与 SVM 几乎相同的检测精度的同时，计算量降低了35倍。

SVM 与 RVM

如前所述，SVM 和 RVM 都是基于核的方法，两者都仅以训练数据的子集为基础作出决策 SVM 中的支持向量和 RVM 中的关联向量，都用来刻画各个类别。但是，SVM 和 RVM 倾向于选择非常不同的向量来表示类别。SVM 选择非常靠近决策边界的向量作为支持向量，而 RVM 则倾向选择两个类别最典型的向量作为关联向量。图4中给出了支持向量和关联向量的实例。请注意， MC 存在 和 MC 不存在 的支持向量非常难于区分，因为它们都位于决策边界的附近，而 MC 存在 和 MC 不存在 的关联向量则分别是病变区域和背景区域的明确实例。

CADx：微钙化点簇的诊断

大量的研究已表明计算机化的 CADx 旨在很难将良性 MC 和恶性 MC 区分时对放射科医师提供帮助。在文献[15]中，证明了一个 CADx 系统能够比放射科医师更准确地分类点簇化的 MC。该方法使用了前馈神经网络（FFNN），使用从点簇化 MC 图像自动提取的测度进行训练。

在机器学习最新进展的推动下，我们在文献[16]进行查找，以便确定最先进的机器学习方法[SVM、核 Fisher 判别（KFD)、RVM、委员会机器（包括总体均值和 Adaboost，这是一个众所周知的boosting方法）]，较之前的方法，如 FFNN，能够进一步改善将 MC 点簇分为恶性或良性肿瘤的分类性能。我们使用了文献[15]中定义的以单个 MC 的形状和大小为依据的特征，以及作为一个点簇的总体分布，已知这些都都与放射科医师常用的特征定性相关。

评价研究表明，核方法（SVM、KFD 和 RVM）性能上彼此类似（就受试者操作特性（ROC）曲线下面的区域而言），但在统计意义上，其性能都比 FFNN 和 AdaBoost 显着改善。

基于内容的图像检索（CBIR）CADx

虽然看起来很途光明，但 CADx 在应用于临床实践时遇到阻力，部分原因是放射科医生在进行培训都是解释视觉数据，很少处理定量的 X 线信息，如乳腺恶性肿瘤的可能性等。因此，当提出一个数值，而没有其它的支持证据，即便是放射科医生也很难完美地将这个数字纳入到诊断结论中。就其本身而论，传统 CADx 分类器经常被质疑为是一个 黑盒子 方法。

为了避免这一缺陷，我们一直主张采用另一种方法，也就是基于内容的图像检索（CBIR）[17][18]，使用一个图像搜索引擎通过展示来自过去案例中的相关信息，在有困难的案例中为放射科医生的诊断听过信息。检索到示例病变能够使放射科医生明确地比较已知病案和未知案例。这种方法的主要优点在于能够提供针对病案的证据，以支持放射科医师作出基于病案的推断，而不是作为一个附加的决策者。

对于一个检索系统而言，要成为一个有用的诊断助手，检索到的图像必须与放射科医师感知到的查询图像真正相关，否则他们可能只是简单地将其忽略。在2000[17]年，我们提出了一个有监督学习方法，用于对放射科医师的图像相似性概念进行建模，以便用于 CBIR。我们的基本原理是对通用的图像检索使用设计的数学上的距离测度，可能不能充分刻画图像图像临床概念上的相关性，这些都是专家观察者所作的复杂评估。

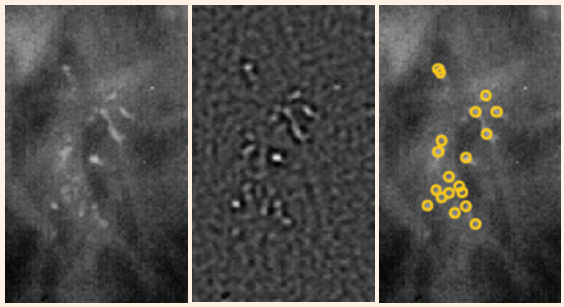
在我们的方法中，通过应用于图像特征的非线性回归模型对两幅病变图像的感知相似度进行建模。该模型通过对实例使用有监督学习而确定，这些实例收集自人类观察者的研究，或者来自在线用户反馈（在该系统使用期间所获得的）。具体来说，我们首先通过包含其关键相关特征的向量u刻画病变。接下来，通过预测模型$f(u,v)$产生一个相似度系数（SC），比对特征向量u和数据库条目的相应特征向量v。具有最高 SC 值的图像将被从数据库中检索出来，并呈现给用户。在我们的研究中，采用了非线性回归 SVM 和一般回归神经网络（GRNN）进行建模$f(u,v)$。事实已证明我们的学习测度远远要比其他可用的度量有效[17][18]。

为了解释说明感知相似度，图6是一个使用多维尺度（MDS）的算法创建的平面图，显示了 30 个微钙化簇。MDS 是一个系列技术，其目的将高维数据映射到低维表示并保持保持相对距离（即如果两个点在高维空间中彼此接近，则 MDS 试图在低维空间中也将它们放置在彼此接近的地方）。

在图6中，散点图中的每个微钙化簇是由一个标记（方形或圆形）表示。MDS 在尝试放置这些点时，使得视觉上类似的微钙化簇（如同人类观察者判断的一样，）在散点图也放置得接近彼此。对应于这些数据点的微钙化簇的实例表示为加号（+）的集合。这些实例的视觉检查表明，平面图的纵轴与微钙化点的密度密切相关，而横轴则反映了点簇的形状。必须要注意的是在这个空间里的恶性和良性病变之间，有一个合理的，但并非十全十美分离平面。

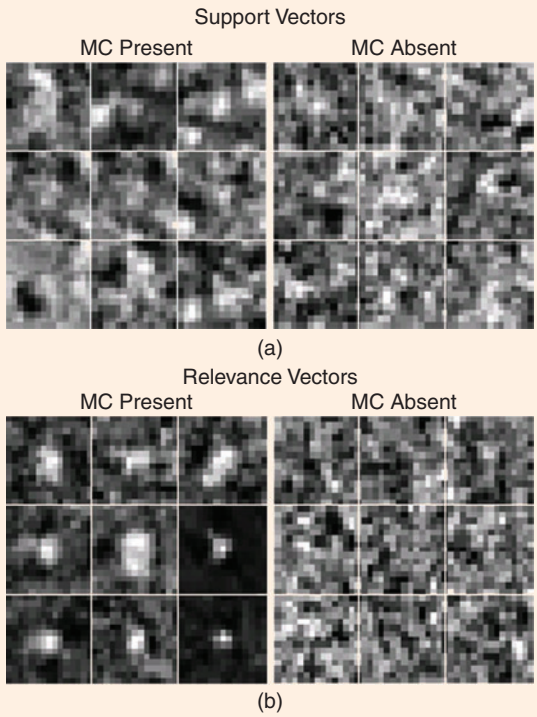
最近，我们建议采用 CBIR 推动传统 CADx 分类器的性能[18]。具体来说，使用与放射科医生评估的图像相类似的数据库图像来改善 SVM 分类器，从而提高在当前病案中的准确度。我们目前正在调研 CBIR 对放射科医师的诊断行为的影响。

诊断行为预测的自动图像质量评估



[图3] (a) 包含微钙化点的乳腺 X 光影响的实例。(b) SVM 检测器的输出 y。(c) 使用 y 后检测到的微钙化点簇位置。

诊断成像可以被看作是由成像设备、图像处理器（例如图像重建算法和显示）和人类观察者（如放射科医生）组成的管道。需要评估图像采集和处理阶段的设计选择对最终解释阶段的影响的原则性方法。



[图4] (a) SVM 的支持向量的比较
(b) 用来检测微钙化点簇的 RVM 的关联向量。SVM 自动选择位于决策边界的实例作为支持向量（因此“MC 不存在”和“MC 存在”的支持向量看起来很类似），而 RVM 则趋于选择的两个类别中更典型的向量作为关联向量（因此两组关联向量看起来非常不同）。

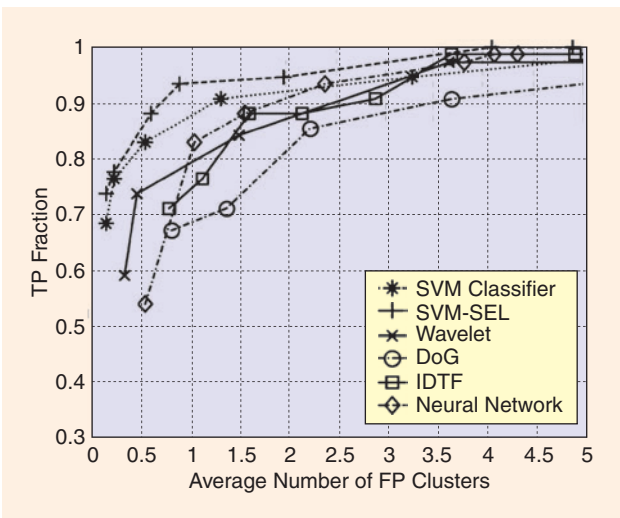
一般传统上评估成像设备和图像重建软件仅仅使用基本的保真度指标，如信噪比（SNR）、均方误差，以及偏置和方差。然而这些指标都是用来比较受统计意义下不同类型的模糊，噪声和伪影影响的图像[19]。在X光成像方面，Lusted 在二十世纪70年代就认识到这一点[20]，他指出，从物理学的观点来看，图像可以忠实地再现组织的形状和纹理，然而并未包含有用的诊断信息。在 Science 上很有影响力的一篇文章中[20]，Lusted 推测，要衡量一个诊断用成像实验的价值，那么在使用成像实验时，我们必须评估观察者的行为。换句话说，如果要将图像用于病变检测，那么图像质量的好坏应由具有检测病变能力的观察者来进行判断。这种方法被成为基于任务的图像质量评价称。

> 尽管重采样没有预测模型那么受关注，但也同样重要。

Lusted 进一步地讲，来自古典检测理论的 ROC 曲线是一种描述诊断行为，进而图像质量的理想的手段。此方法使得 ROC 分析在医学成像领域得到了广泛的应用和实现，例如在 Metz 等人所发布的 Rockit 中[21]。

图 7 给出了人类观察者的行为如何受所呈现的图像类型影响的例子。在这种情况下，呈现给人类观察者的是一幅通过单光子发射计算机断层成像技术（SPECT）获得的心肌灌注影像（心壁）。以使用相同的数据集合通过不同的方式重建的图像为基础，要求该观察者判断是否有指示灌注不足的暗区。图 7 给出了 12 种不同的重建结果，通过使用一步或五步迭代的有序子集期望最大化算法（OS-EM），以及具有可变半高宽（FWHM）的高斯滤波器获得。

病灶位于箭头所指的位置，沿着图7的顶部和底部的数值是观察者所说的信心度（分为1-6个等级，6 表示信心度最高）。请注意，随着图片被处理得越来越光滑，观察者的对于存在病灶的信心度先上升，后下降。对于最佳平滑级别的选择就是目标的一个实例，其中定量图像质量度量是必要的。



[图5] 用于检测乳腺 X 光影像中的 MC 的不同方法的检测性能。通过一个连续学习 SVM 获得了最佳性能，以每幅图像执行一次 FP 聚类的代价即可实现约 94% 的检出率（TP分数），而经典技术（DoG）大约只有 68% 的检出率。

人类观察者的机器学习模型

在诊断成像中，衡量图像质量的黄金标准就是对使用给定图像集合时，测量观察者（如放射科医生）的诊断水平进行统计研究。不幸的是，这些研究费用和复杂性让它们无法日常使用。因此，数字观察者 模仿人类观察者行为的算法目前已被广泛地用作人类观察者的代替物。

一个被称为 Hotelling observer（CHO）[22]的特殊数字观察者，现已被广泛使用，特别是在核医学成像方面。CHO 是一个 Fisher LD，应用于对图像进行带通滤波（通道）而得到的输入特征。这些通道的灵感来自人类视觉系统中有关感受野的概念。由于它对于图像质量评价的原则性方法，CHO 理所当然地在该领域占有非常重要的一席之地，并享有巨大声望。

然而，CHO 并不能完美地捕捉人类观察者的表现，因此，我们提出了一个新的方法，其中基于任务的图像质量评估问题被当做有监督学习或系统识别问题[23]。也就是说，其目的就是确定 x 中的图像特征与观察者评分 y 之间的未知的人类观察者映射 $f(x)$，其中观察者评分反映了人类观察者认为图像中存在异常的信心。可以从来自人类观察者的实例数据学习到该关系，在没有人类观察者数据可用的场合，该模型则可用于在新的情境下做出预测。

在我们的工作中，到目前为止，一直保留 CHO 中使用的通道，均包含在向量 x 中，但我们把这些通道作为 SVM $f(x)$ 的输入，我们将其训练用来在训练实例 $(x_i\ y_i), i = 1,...,N$ 的基础上预测观察者评分。由此产生的算法被称为通道化SVM (CSVM).

结果分析

在文献[23]中，我们比较了 CSVM 和 CHO 在心肌 SPECT 成像中对图像质量的评价。在这个实验中，在一个涉及到其他60个图像的培训课程后，两个医用物理学家评价了100幅嘈杂图像中的缺陷可见度，并就确定一个6个点大小的病灶的信心进行评分。人类观察者在平滑滤波器有6个不同的选择，OS-EM 重构算法的迭代次数也有两个不同选择的情况完成此任务（见图7）。

为了证明这一方法的泛化能力，我们用分布广泛的图像训练了 CHO 和 CSVM，然后都在一个不同的，但同样也很广泛的图像范围内进行测试。具体来说，我们使用滤波器 FWHW 的每个值和 OS-EM 五次迭代下的图像训练了两个观察者，然后使用滤波器 FWHW 的每个值和 OS-EM 一次迭代下的图像对观察者进行测试。仅使用基于训练图像的五折交叉验证对 CHO 和 CSVM 的参数进行了充分优化，以便将测量泛化误差降到最小。因此，对于两个观察者而言，测试图像没有以任何方式用于模型参数的选择。图8中对位于 ROC 曲线下面（AUC）的数字观察者区域的数字观察者的预测进行了比较，以便反映人类观察者的真实能力。在这种情况下，CHO 表现相对较差的，无法匹配人类观察者 AUC 曲线的形状或幅度，而 CSVM 能够在这两种情况下生成相当准确的 AUC 预测。每个错误条都代表对测试数据进行 5 折交叉验证时所计算的标准偏差。

这个实验证明了利用机器学习，而不是固定模型生成预测有着潜在的好处。由于该方法的普遍性，机器学习可以用

来对人类观察者在许多临床工作中的行为作出预测，而不只是病变检测，而 CHO 是专门设计用于病变检测，因此不适合进行推广。

> 对于一个检索系统而言，要成为一个有用的诊断助手，检索到的图像必须与放射科医师感知到的查询图像真正相关，否则他们可能只是简单地将其忽略。

脑功能定位

脑成像涉及到对大脑空间表征（图）的创建，有助于了解处于正常和疾病过程中各脑区的作用。脑成像是其中一个非常不同的应用领域，因此，我们只针对以下两个主要方面进行了讨论：1）在许多情况下，脑成像对预测输出 y 的关注要远远低于从脑图中获得的模型 $f(x)$ 本身；以及 2）由于脑成像中可用的数据实例数量相对较少，非线性模型并不总是优于简单的线性方法。

脑成像领域至少已经快速增长了 25 年。在本文有限的篇幅内是不可能给出本领域的均衡调查及其使用的机器学习，所以我们只给出一个简要概述。

在20世纪80年代，主导脑成像的技术是正电子发射断层扫描（PET）和SPECT。脑功能图分析中的第一个机器学习方法就是将人工神经网络（ANN）应用于糖代谢的 PET 图像[24]。然而，随着在1990年发现血氧水平依赖（BOLD）的信号，可以间接测量区域神经活动，功能磁共振成像（fMRI）和相关技术的应用有了爆炸性增长[25]。

当时脑成像中的实验和分析范式仍是以单变量一般线性模型（GLM）与推论统计检验[26]为基础，在某些情况下则是以他们的预测，机器学习当量，高斯简单贝叶斯[27]为基础。使用相关的多元分类方法的论文近期有所增多，这被该领域的某些人称为 读心术 。最近的综述，包括历史的观点，参见文献[28]，往往都忽略简单的多元方法，如主成分分析和 LD，应用于疾病群的 PET 扫描[29]，其中反映了针对大脑网络的测量协方差结构，超过20年的工作成果。在最近的 fMRI 脑成像文献中，这一网络主题最近势头强劲，主要聚焦于测绘所谓的 默认模式 脑网络，使用的手段包括逐对体素相关性[30]，或种子像素/行为的偏最小二乘法（PLS）[31]，独立成分分析（ICA）[32][33]，以及最近的非线性动力学[34]和图论与白的问题，再加上脑白质网络的结构化扫描[35]。

我们自己的大部分工作集中在如何评价和优化性能的问题上，以及如何从广泛的可用机器学习工具挑选最佳的信号探测器。我们尤其致力于研究较小的样本规模的影响，其中将渐近分析理论应用于多元机器学习模式，如果影响存在，则无法提供太多，如果有的话，则用于引导。脑图分析是一个非常不适定的问题，其中通常有数万或数十万计的体素，但只有数十或数百次大脑扫描。因此，小样本的限制对于脑成像中的医学用途而言，可能是最重要的。

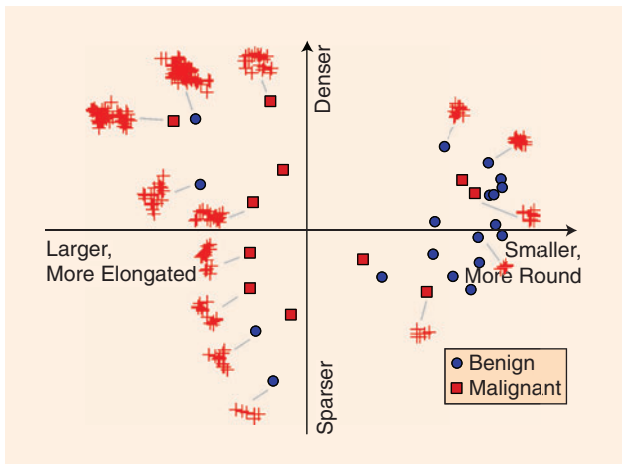
判别脑图

为了说明机器学习在脑成像中的应用，让我们设想这样的研究，在其中我们希望产生一个图像用来显示新药对脑功能的区域性影响（本文的两位的作者，Wernick 和 Strother，就是针对制药业进行这样的商业分析）。要做到这一点，可以对 N 位受试者组成的小组进行两次扫描，一次是在服用新的药物后，一次是在投用安慰剂后。然后分析 2N 幅图像，以获得描述药效的图像。我们希望这一发现不仅能描述这个特定的受试者小组，也能推广到更广泛的人群。

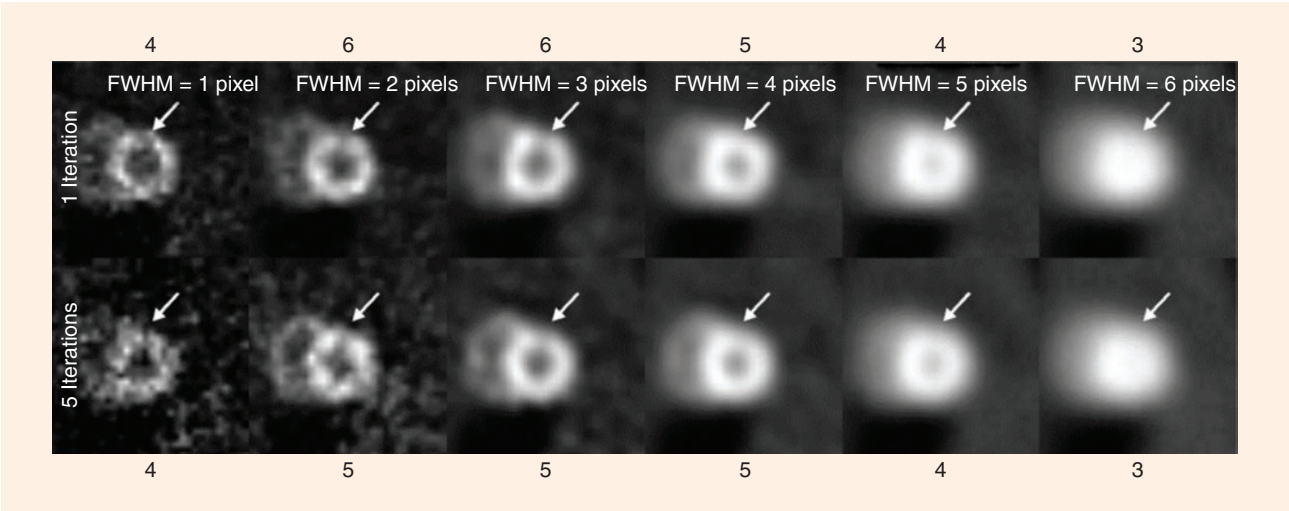
针对这个问题，许多机器学习方法潜在的基本思想就是在高维空间中将每个图像当作一个向量，每个分量都代表扫描中的一个体素的值。在这个例子中，我们的数据可以被看作是由两类图像组成：药物和安慰剂。为了将维数降低到可管理的水平，并降低噪声，常见的方法是通过奇异值分解（SVD）变换数据。接下来，一个训练过的分类器根据降维数据将药物图像和安慰剂图像区分开来。

在传统的模式分类应用中，训练分类器的目的是要对新数据作出决策。事实上，神经影像学中这样的例子越来越多，例如在测谎，或个别病人的疾病诊断中。然而，在许多研究中，其目的仅仅是为了了解大脑在不同的条件下本质上有什么不同，也就是说，药物和安慰剂条件。在这种情况下，所需的信息就编码在预测模型 $f(x)$ 自身中。当采用线性模型时，期望的脑图被编码在判别向量中，而判别向量（从 SVD 空间反投影到图像空间）是用来描述物和安慰剂条件下体素的突显度。

图 9 给出了这样一个图像的例子（我们称其为空间激活模式），在经过阈值处理后，叠加在一个用于将多个受试者的大脑移动到大致相同空间的模板结构化图像上。此图像中每个染色的体素表示体素对药物与安慰剂区别的程度，因此该图像描绘了影响的空间分布。



[图6] 可视化各个乳腺X线影像上出现的异常间关系的统计工具，其中距离反映了异常间相对的相似之处，则由人类专家进行判断。通过多维尺度分析，一个力求在可以随时可视化的较低维图上表示高维数据，同时又能保持数据点间相对距离（相似度）的统计工具，将 MC 点簇描绘在这个两维图上。每个红色加号（+）组描绘了与散点图中给定的点相关的真实 MC 点簇。这表明所绘图的纵轴与各点簇的密度相关，而横轴则是与其形状相关。



[图7] 一个人类观察者有关出现异常的判断（在心肌灌注缺损的情况下），是依据用于创建图像的重构算法参数给出的（此处的参数为迭代次数，后重构光滑核的宽度（FWHM））。 以上所有图像的缺损位置都用箭头作出指示，但任何人士要求来判断是否有不同意见，在他们从三个值的缺陷，即“缺陷可能是不存在，”到六个价值，意思是“缺陷确实存在。”我们的算法预测这种行为的能力使得我们能够为一个具体的诊断任务优化给定的算法。

请注意，在这个基础介绍中，我们一直都没有描述在使用机器学习算法之前需要应用的一系列重要的预处理步骤。文献[36]中进行了详细的讨论。

模型、样本大小和信噪比

对数据分析技术的评价已清楚地表明，最佳工具的选择关键取决于手头数据的信号和噪声结构，样本大小[37][38]。例如，图10（来自文献[38]）就说明了直到有足够的数据实例支持对非线性模型中固有的较多参数的估计之前（在这种情况下是 ANN），简单的线性模型优于一个灵活的非线性模型然而，在目前这些脑成像的文献中，讨论或比较不同的分析技术时，这些问题常常被忽略。

我们已经讨论了使用文献[39]中的仿真方法，基于图 11 中所给出的简单仿真模型选择最优的分析过程，假设实验设计类似于前面所述的药物 安慰剂研究。我们修改了仿真实验中的多个参数，包括每一个环境中的实例数码（从20 到 100），以及仿真模型中激活 斑点 的振幅（基准线的 3 % 或 5% 以上）。我们添加了空间上有色，时间上为白色，标准差为平均基准值 5% 的高斯噪声。我们创建了斑点的三个空间分布 网络 ，并改变它们之间的相关系数ρ（rho）（ρ= 0.0、0.5 或 0.9)）以及他们的幅度方差和噪声方差的比率V。该比率可以认为与音频的动态范围相类似，因为在本应用中，斑点方差也是信号的来源之一，这与该领域最近聚焦于脑成像中的网络监测特别相关。在文献[39]中，我们发现 SVD 本身或再加一个 LD 更适合于子空间，在其上对于网络互动的估计敏感度要高于逐对相关系数[40]。

我们使用相同的仿真模型重复并延伸 Lukic 等人先前的工作（结果如图12所示果）。仿真包括3%的高斯振幅，有 30 个基准线和 30 激活扫描。所测试的模型包括：1）使用局部（GLM-S）和空间（GLM-P）方差估计的单体素测试，分类计数包括2）两类 Fisher LD，3）归一化 LD（NLD），以及4）二次判别函数（QD）。所有的多元技术都是在 SVD 空间上进行估计，并利用贝叶斯证据优化确定维数[41]，就如在软件包 MELODIC 所估计的那样[42]。对于 LD 和 QD 而言，SVD 基本分量的长度与其特征值相等，对于 NLD 来说，则将其归一化到单位长度。

使用 ROC 曲线下方的区域，其误报率在[0.0，0.1]之间，信号检测通过横跨高斯斑点峰值处的 16 个体素进行测量。即便使用局部方差估计（GLM-S）的t检验是 正确的模型（即$V=0.1$），将t检验和合并方差估计或自适应多元协方差检测器一起使用可获得更好的检测性能。此外，GLM-S 的性能明显下降，因为随着V的增加等方差假设不再成立。通过空间合并（GLM-P）的方差估计则显着改善了信号检测，并大大地清除了违规模型的来源。

通过 LD 的结果给出了 GLM-S 违规模型的多元等价式，随着V的增加违反了类内协方差相等（例如，基线和激活扫描使用同一个网络结构）的假设；只有激活扫描有一个非对角线的类内协方差结构，会随着V的增加而增加。但是，除了在 rho 和V很大，严重违反相等协方差假设时，LD 仍优于 GLM-S，如图12（c）所示。在 NLD 方法中，归一化输入特征方差（即单位化 SVD 偏差向量）的标准机器学习方法可以显著改善信号检测性能，通常都优于 GLM-P，且随着 V 的增加，大大地消除了 LD 性能的下降。最后，在假设有不同类内协方差的情况下，使用正确的多元模型，QD 可以进一步改善性能，直至接近于完美（部分 ROC 曲线接近于 0.1）这里用到了 QD，可作为针对非等类分布的问题，SVM 的替代解决方案，如图2所示。

在脑成像中，LD 类方法和 SVM 的相对性能仍未达到共识，某些文章声称 SVM 很优越[43]，其他的则基本都相差无几[44]，但它们对于不同输入 SNR 结构的相应有所不同，如图 2 中的分析。此外，我们最近的仿真结果表明，信号检测性能是基集合大小的强函数，性能可能会进一步改善，甚至比图11中所示的基于如下所述的重复性度量的最优 SVD 子空间的重采样估计更好。

我们最后的仿真结果比对了利用贝叶斯核方法与广义似然比检测对功能神经成像中的局部激活进行估计的结果。在文献[45]中，我们比较了使用空间高斯核叠加的信号检测（其参数通过基于可逆跳马尔可夫链蒙特卡罗（RJMCMC）算法的最大后验（MAP）技术从数据中估计而得）和 RVM 方法。RVM 和 RJMCMC 是比文献[39]中所尝试的其它所有方法都更优的信号探测器，在 ROC 曲线一下的部分区域，可以达到 0.80 和 0.82。这些性能值不能直接与图 11 相比较，因为其仿真参数完全不同。然而，即使是在我们简单的体模中，RJMCMC 需要数十小时进行计算，而 RVM 仅需计算数分钟而已。脑成像中的 SVM、RVM 和其他核技术（例如核PCA[28]、核典型相关分析[46])仍有待创立。

数据驱动的性能指标

在脑成像中，作为一般机器学习的应用，优化和评估预测模型，并选择其最显著的特征是非常重要的。这些任务必须遵循一个定量的性能指标。预测精度经常扮演这一角色，例如指导贪婪搜索过程来选择最突出的体素集[26]。在文献[4]和[27]中对这类纯粹的预测驱动分析方法中的一些需要权衡的问题进行了讨论。

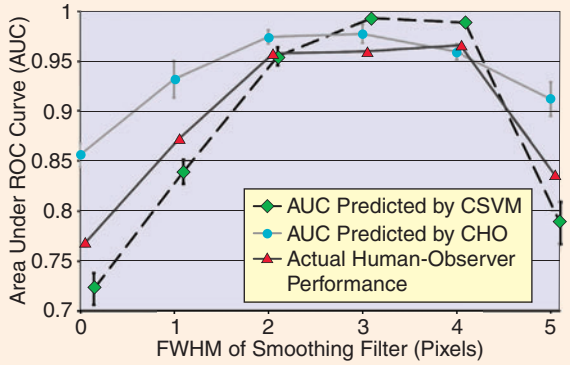
尽管预测精度可以单独作为一般机器学习问题的有效度量，神经影像学也要求空间模式（通过预测模型编码）可以在不同的受试者群体或同一受试者的不同扫描间可重复。和预测精度一起，重复性一项重要的度量，这是一个非常有效的数据驱动的 ROC 分析的替代品。

Strother 等提出了一种被称为 NPAIRS 的新颖的折半重采样的框架[9]，同时可以评估其预测精度和重复性。可达到的预测精度和模型重复性间的权衡涉及到估计论论中经典的偏差折中法。在此应用中，预测精度的获取一般都以降低重复性为代价，反之亦然。通过绘制预测精度与重复性曲线，作为某些参数的函数（如 SVD 基向量的个数），我们可以评估整个范围内的折中效果，与 ROC 曲线，信息检索领域的查准率　查全率曲线曲线场，或来自统计的偏差-方差曲线极为相似。我们将 NPAIRS 分析所生成的曲线称为 (p,r) 曲线。

为了使用 NPAIRS 计算(p,r)曲线，将该数据集的独立观测量（如跨越受试者）进行等分为：训练和测试集合。在其中一个半分集合内应用空间模式获取预测精度（即训练）以便在另外一个半分集合（即测试）中估计扫描类标号。然后将两个半分集的角色互换，也就是说每个集合都有一次用作训练集合（为了生成空间激活模式），有一次用作测试集合。从这些结果来看，计算的两种预测精度的估计值估算(p)并平均，以便获得整体的预测精度。接下来，计算两个独立空间激活模式的重复度，用作两种模式中空间所有的定位体素对的关联度（r）。此关联度r直接关系到每个半分模式的提取对中的可用 SNR。如果其中一个来自一个空间模式的体素值所形成的散点图，另一个则相应地来自另一个空间模式，其中一个就会得到这样的分布，相关性很较强，或者说信号，轴的相关特征值为（1+r），非相关性较小，或者是噪声，轴的特征值为（1-r）。因此，可以定义一个全局数据信噪比度量 gSNR 为：

$$gSNR = \sqrt{((1+r)-(1-r))/(1-r)} = \sqrt{2r/(1-r)}$$

.



[图8] 通过机器学习方法（CSVM）预测人类观察者的性能（AUC）与传统的数字观察者（CHO）相比较。CHO 并不没有认识到在较低和较高水平平滑会降低诊断性能的程度，沿着图 7 的顶部和底部的数值就可以看到分数图7看到影响。

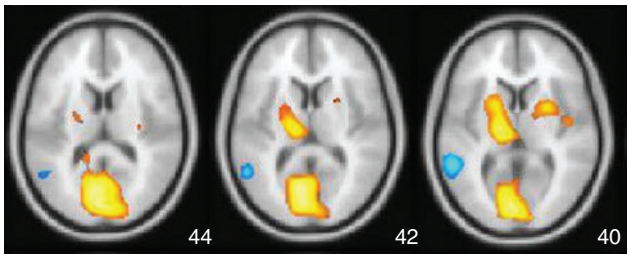
在 NPAIRS 中，多次执行半分重采样，然后对结果平均或取中位数，并记录其分布特征。这种重采样的方法对于通过 0.632+ bootstrap 方法获得平稳鲁棒度量非常有好处。最后，使用得到大家一致共识的技术将多个半分空间模式组合成单一的以 Z-score（标准正态分布）量表描述的模式，对于可产生基于体素的参数估计值的所有预测模型提供稳健的 Z-score 机制。

在文献[29]中，NPAIRS 被应用于PET，同样也被应用于 fMRI [47]–[49]。尽管 NPAIRS 可应用于任何分析模型，但我们还是特别聚焦于 LD 方法，以及新近出现的 QD 法，这两者都是建立于 SVD 的基之上。 这使我们能够 1）在 SVD 的基础上，通过选择的软阈值（如脊形）或硬阈值或其他基集合对模型正则化[50]，2）维持协方差分解链路，已证明在 PET 中对于阐明网络结构非常有用，以及 3）生成全脑激活图，可提高发现脑功能和疾病的新特征的可能性。

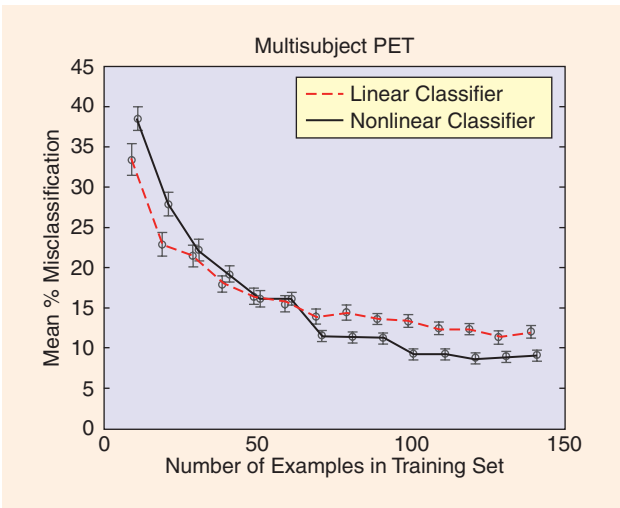
图 13 给出了一个实例，说明如何将 NPAIRS 用于研究图像分析过程中关键参数的影响，从而实现对这些参数的最佳选择。在这个例子中，分析了 fMRI 图像分析过程中的两个参数：SVD 基向量的个数（定义模型复杂度）和用于趋势剔除的半余弦的数目[36]。我们不会在这里详述 SVD 和趋势剔除工具的技术细节；我们给出这个例子仅仅是要说明一般情况下，如何使用使用 NPAIRS 选择最佳的模型参数。）

在一幅(p,r)图中，通过到达空间的右上角，在那里预测准确度（如图 13 中所描述的后验概率）达到1.0，重复性也达到1.0，即可获得理想的性能 。因此，确定参数的最佳的方法就是确定这样一个点，在该点(p,r)曲线与点（1,1）间的欧氏距离（$\overline{M}$）最小。在这个例子中，我们可以看到，随着 SVD 分量的增加，性能[到（1,1）的距离]有改善，然后再恶化。余弦去趋势参数的影响较弱，但指出一个半周期相对两个周期而言是更好的选择。在这个图中，五至十个 SVD 分量之间的钩状部分代表了 fMRI 中普遍的重复性伪影。

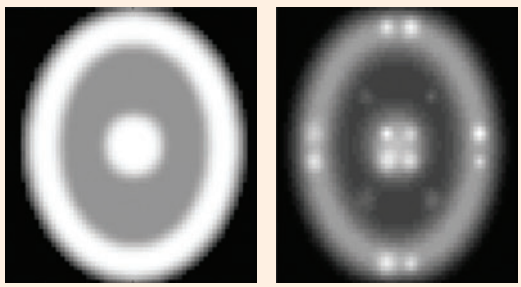
在有关脑成像的挑战性问题中，NPAIRS 分析框架为理解和优化模型性能提供了非常有用的方法，或许也可以用于



[图9] 大脑中的空间激活模式，显示了抗焦虑/抗抑郁药物丁螺环酮（Buspar）的影响，通过在 12 位受试者的 FDG-PET 影像上应用 Fisher LD 和 NPAIRS 半分重采样而获得（数据由 Abiant, Inc. 提供；由 Predictek, Inc. 分析）.结果表明纹状体活化（上部橙色区域）可能是因为该药物作为多巴胺 D2 受体拮抗剂的行为之一。



[图10] 这些交叉的学习曲线（分类器的性能与训练集的大小）表明，当训练样本数量比较少时，也可以通过一个简单的多元线性分类器（这里为 Fisher 判定）构成非线性分类器（本例中为神经网络）。这并不是不可预见，因为小数据集一般不支持复杂的模型，但这一结果强调了研究人员要抵制在各种情况下使用高复杂性模型的诱惑的重要性。



[图11] 用于测试信号检测性能的仿真体模。

其他应用，其中不仅对精确预测感兴趣，还对生成驱动这些预测的有关因素的可靠信息颇有兴趣。

致谢

作者们希望对本文中所总结的研究工作作出贡献的众多合作者表达真诚的谢意，其中有 Nikolas P. Galatsanos ，Lars Kai Hansen ，Issam El-Naqa, Ana S. Lukic ，Robert M. Nishikawa，Stephen LaConte，David Rottenberg，Liyang Wei 和 Jane Zhang。

本文中所回顾的研究工作由 NIH/NCI （CA89668）、NIH/NIBIB（R01EB009905）、NIH/NIBIB （HL091017）、NIH/NINDS（NS34069）、NIH/NIMH（MH073204）、James S. McDonnell 基金会为 Grigori Yourganov 所提供的博士奖学金 Ph.D. scholarship to , NIH/NIBIB P20EB02013, NIH/NIMH P20MH072580, and CIHR/MOP84483 资助。Stephen C. Strother 要特别感谢安大略省心脏与中风基金会通过中风康复中心所给予的支持。

作者简介

Miles N. Wernick（wernick@iit.edu）于1983年在西北大学获得物理学学士学位，1990年在罗切斯特大学获得光学博士学位。1990年，作为美国国立卫生研究院博士后研究员在芝加哥大学研究放射学，在那里他成为一个副研究员助理教授。1994年，他加入了伊利诺伊理工大学，目前他是医学影像研究中心的主管，电气和计算机工程以及生物医学工程系的Motorola基金会的讲座教授。他还是 Predictek 公司的总裁，他的研究兴趣包括医疗成像，机器学习，图像处理，光学。他是IEEE Signal Processing Magazine特刊的客座编辑，IEEE Transactions on Image Processing和SPIE/IS&T Journal of Electronic Imaging的副主编，IEEE Bioimaging and Signal Processing Technical Council的会员。

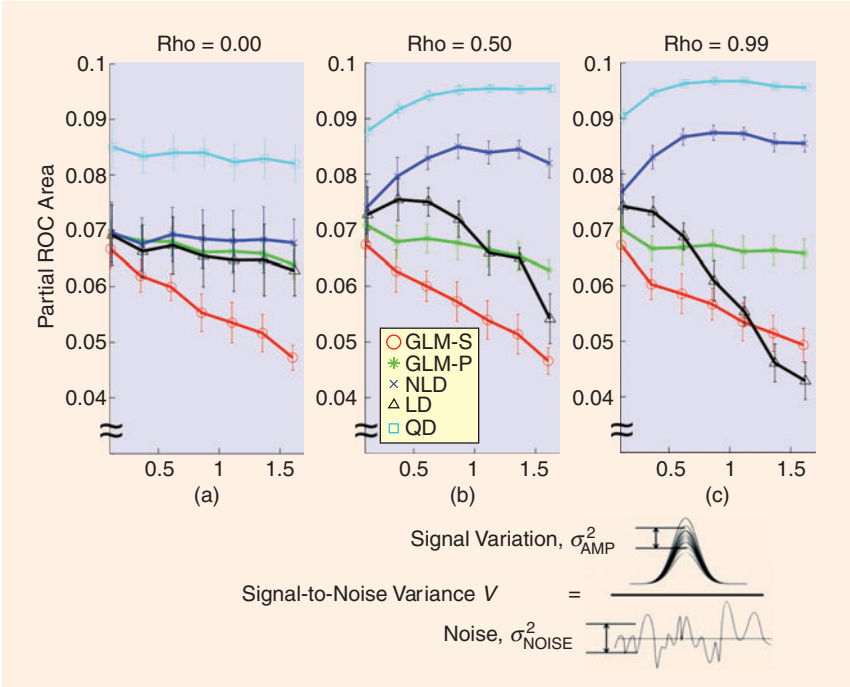
Yongyi Yang (yy@ece.iit.edu) 分别于1985年和1988在中国北京的北方交通大学获得电子工程学生学位和硕士学位。并分别于1992年与1994年芝加哥的伊利诺伊州科技大学（IIT）获得应用数学硕士学位和电气工程博士学位。他目前是IIT在电气与计算机工程系任教授，他在医学影像研究中心工作，而且是生物医学工程学系的合聘教授。他的研究兴趣包括信号和图像处理、医学成像、机械学习、模式识别和生物医学应用。他同时还是IEEE Transactions on Image Processing的副主编。

Jovan G. Brankov（ brankov@iit.edu）于1996年在南斯拉夫贝尔格莱德大学获得电气工程高级文凭。并分别于1999年和2002年从伊利诺伊州科技大学（IIT）获得电子工程的硕士和博士学位。他目前在IIT的电气与计算机工程系担任助理教授，他在医学影像研究中心工作。他的研究兴趣包括医学成像、图像序列处理，、模式识别和数据挖掘。他目前的研究课题包括医学图像序列的四维和五维断层图像重建，多图统计法（一种相位敏感的成像方法），以及基于人类观察者模型的图像质量评估。他已编写，或与他人合著超过八十的著作，并担任 Medical Physics 的专案副主编。

Grigori Yourganov (gyourganov@rotman-baycrest.on.ca）分别于2000年和2005年获得位于加拿大多伦多的约克大学的计算机科学的学士学位和硕士学位。他目前正在在Rotman研究所（多伦多大学）医学科学研究所攻读博士学位，师从于 Stephen C. Strother 博士和 Randy McIntosh 博士。他的研究

主要集中在将多变量分析技术用于fMRI数据的应用。

Stephen C. Strother（sstrother@rotman-baycrest.on.ca）分别于1976年和1979年在位于新西兰的奥克兰大学获得学士学位和硕士学位，1986年在蒙特利尔的McGill大学获得电器工程的博士学位。Since 1985, he has been a postdoctoral fellow at Memorial Sloan-Kettering Cancer Center, New York.在1989年，他作为高级 PET 物理学家加入了位于 Minneapolis 的弗吉尼亚州医学中心，并在2002年成为明尼苏达大学的放射学教授。2004年他移居到多伦多，作为 Rotman Research Institute 的资深科学家和多伦多大学的医用生物物理学教授，在那里他还是中风康复中心的是 multi-institutional 中风康复中心的核心成员。目前的研究兴趣包括神经信息学，主要致力于将机器和统计学习技术用于 PET 和 fMRI/MRI 神经影像在大脑老化的研究与临床应用中的应用。 2001年他与他人一起在芝加哥创办了 Predictek 公司。他还是Human Brain Mapping的副主编。



[图12] 在(a)－(c)中给出了作为被激活大脑区域的网络间的信噪方差比（V）和相关性（rho）的函数，五种模式下检测大脑激活的性能。QD 和 NLD 的表现最佳，随着网络强度的增加而有所改善（增大了 V 和 Rho 值），而单变量方法的性能已经落后，实际上随着信号强度的增加而恶化。

参考文献

[1] T. Hastie, R. Tibshirani, and J. H. Friedman, The Elements of Statistical Learning.New York:Springer-Verlag, 2003.

[2] B. Scholkopf and A. J. Smola, Learning with Kernels:Support Vector Machines, Regularization, Optimization, and Beyond.Cambridge, MA:MIT Press, 2001, p. 626.

[3] M. N. Wernick, Pattern classification by convex analysis, J. Opt.Soc. Amer.A, Opt.Image Sci., vol. 8, pp. 1874–1880, 1991.

[4] V. N. Vapnik, Statistical Learning Theory.New York:Wiley, 1998.

[5] M. E. Tipping, Sparse Bayesian learning and the relevance vector machine, J. Mach.Learn.Res., vol. 1, pp. 211–244, Sept. 2001.

[6] R. G. Baraniuk, E. J. Candès, R. Nowak, and M. Vitterli, Compressive sampling, IEEE Signal Processing Mag., vol. 21, no. 2, pp. 12–13, Mar. 2008.

[7] B. Efron and R. J. Tibshirani, An Introduction to the Bootstrap.Boca Raton, FL:CRC, 1994.

[8] B. Efron and R. Tibshirani, Improvements on cross-validation:The .632+ bootstrap method, J. Amer.Statist.Assoc., vol. 92, no. 438, pp. 548–560, June 1997.

[9] S. C. Strother, J. Anderson, L. K. Hansen, U. Kjems, R. Kustra, J. Sidtis, S. Frutiger, S. Muley, S. LaConte, and D. Rottenberg, The quantitative evaluation of functional neuroimaging experiments:The NPAIRS data analysis framework, Neuroimage, vol. 15, no. 4, pp. 747–771, Apr. 2002.

[10] Image-Processing Techniques for Tumor Detection.New York:Marcel Dekker, 2002.

[11] Recent Advances in Breast Imaging, Mammography, and Computer-Aided Diagnosis of Breast Cancer.Bellingham, WA:SPIE, 2006.

[12] J. Tang, R. M. Rangayyan, J. Xu, I. El Naqa, and Y. Yang, Computer-aided detection and diagnosis of breast cancer with mammography: recent advances, IEEE Trans.Inform.Technol.Biomed., vol. 13, no. 2, pp. 236–251, Mar. 2009.

[13] I. El-Naqa, Y. Yang, M. N. Wernick, N. P. Galatsanos, and R. M. Nishikawa, A support vector machine approach for detection of microcalcifications, IEEE Trans.Med.Imaging, vol. 21, no. 12, pp. 1552–1563, Dec. 2002.

[14] L. Wei, Y. Yang, R. M. Nishikawa, M. N. Wernick, and A. Edwards, Relevance vector machine for automatic detection of clustered microcalcifications, IEEE Trans.Med.Imaging, vol. 24, no. 10, pp. 1278–1285, Oct. 2005.

[15] Y. Jiang, R. M. Nishikawa, D. E. Wolverton, C. E. Metz, M. L. Giger, R. A. Schmidt, C. J. Vyborny, and K. Doi, Malignant and benign clustered microcalcifications: automated feature analysis and classification, Radiology, vol. 198, no. 3, pp. 671–678, Mar. 1996.

[16] L. Wei, Y. Yang, R. M. Nishikawa and Y. Jiang, A study on several machine-learning methods for classification of malignant and benign clustered microcalcifications, IEEE Trans.Med.Imaging, vol. 24, no. 3, pp. 371–380, Mar. 2005.

[17] I. El-Naqa, Y. Yang, N. P. Galatsanos, R. M. Nishikawa, and M. N. Wernick, A similarity learning approach to content-based image retrieval: application to digital mammography, IEEE Trans.Med.Imaging, vol. 23, no. 10, pp. 1233–1244, Oct. 2004.

[18] L. Y. Wei, Y. Y. Yang, M. N. Wernick, and R. M. Nishikawa, Learning of perceptual similarity from expert readers for mammogram retrieval, IEEE J. Select.Topics Signal Processing, vol. 3, no. 1, pp. 53–61, Feb. 2009.

[19] N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, and A. C. Bovik, Image quality assessment based on a degradation model, IEEE Trans.Image Processing, vol. 9, no. 4, pp. 636–650, Apr. 2000.

[20] L. B. Lusted, Signal detectability and medical decision making, Science, vol. 171, pp. 1217–1219, 1971.

[21] C. E. Metz, B. A. Herman, and J. H. Shen, Maximum-likelihood estimation of ROC curves from continuously-distributed data, Stat.Med., vol. 17, no. 9, pp. 1033–1053, 1998.

[22] K. J. Myers and H. H. Barrett, Addition of a channel mechanism to the ideal-observer model, J. Opt.Soc.Amer.A, Opt.Image Sci., vol. 4, no. 12, pp. 2447–2457, Dec. 1987.

[23] J. G. Brankov, Y. Yang, L. Wei, I. El Naqa, and M. N. Wernick, Learning a channelized observer for image quality assessment, IEEE Trans.Med.Imaging, vol. 28, no. 7, pp. 991–999, July 2009.

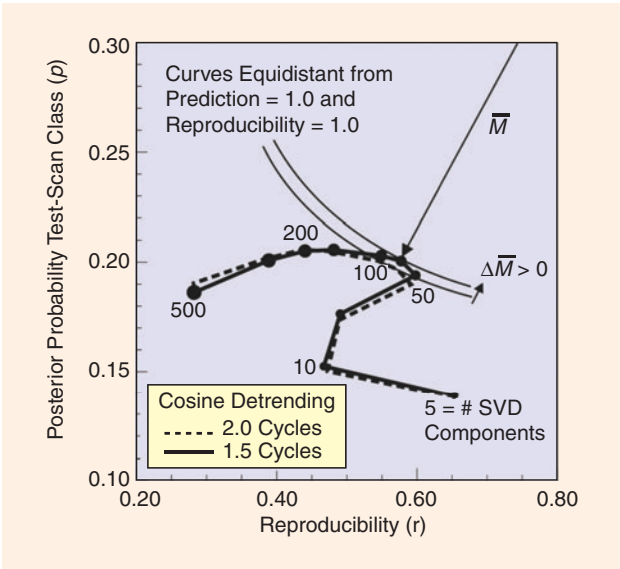
[24] J. Kippenhan, W. Barker, S. Pascal, J. Nagel, amd R. Duara, Evaluation of a neural network classifier for PET scans of normal and Alzheimers disease subjects, J. Nucl.Med., vol. 33, pp. 1459–1467, 1992.

[25] P. Bandettini, Functional MRI today, Int. J. Psychophysiol., vol. 63, no. 2, pp. 138–145, Feb. 2007.

[26] K. J. Friston, J. T. Ashburner, S. J. Kiebel, and T. E. Nichols, Statistical Parametric Mapping:The Analysis of Functional Brain Images.New York:Academic, 2006.

[27] F. Pereira, T. Mitchell, and M. Botvinick, Machine learning classifiers and fMRI:A tutorial overview, Neuroimage, vol. 45, no. 1 (Suppl.), pp. S199–S209, Mar. 2009.

[28] L. K. Hansen, Multivariate strategies in functional magnetic resonance imaging, Brain Lang., vol. 102, no. 2, pp. 186–191, Aug. 2007.



[图13] 在 NPAIRS 框架中，预测—重复性(p,r)曲线给出了预测准确度（纵轴）和由此产生的脑图（横轴）重复性的折中。当曲线最接近理想点（1, 1），实现了最小距离$\overline{M}$时就可实现最佳性能。这为优化图像分析过程提供了基础，本例是在特定 fMRI 数据分析问题中指定了最佳参数（SVD 分量的数目，以及在特定余弦去趋势步骤中的周期数目）

[29] D. Eidelberg, Metabolic brain networks in neurodegenerative disorders:A functional imaging approach, Trends Neurosci., vol. 32, no. 10, pp. 548–557, Oct. 2009.

[30] M. D. Fox and M. E. Raichle, Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging, Nat. Rev. Neurosci., vol. 8, no. 9, pp. 700–711, Sept. 2007.

[31] A. R. McIntosh, W. K. Chau, and A. B. Protzner, Spatiotemporal analysis of event-related fMRI data using partial least squares, Neuroimage, vol. 23, no. 2, pp. 764–775, Oct. 2004.

[32] C. F. Beckmann, M. DeLuca, J. T. Devlin, and S. M. Smith, Investigations into resting-state connectivity using independent component analysis, Philos. Trans.R. Soc.Lond.B Biol.Sci., vol. 360, no. 1457, pp. 1001–1013, May 2005.

[33] N. M. Correa, T. Adalı, Y.-O. Li, and V. D. Calhoun, Canonical correlation analysis for data fusion and group inferences, IEEE Signal Processing Mag., vol. 27, no. 4, pp. 39–50, 2010.

[34] K. E. Stephan, L. M. Harrison, S. J. Kiebel, O. David, W. D. Penny, and K. J. Friston, Dynamic causal models of neural system dynamics:Current state and future extensions, J. Biosci., vol. 32, no. 1, pp. 129–144, Jan. 2007.

[35] C. J. Honey, O. Sporns, L. Cammoun, X. Gigandet, J. P. Thiran, R. Meuli, and P. Hagmann, Predicting human resting-state functional connectivity from structural connectivity, Proc.Nat. Acad.Sci.USA, vol. 106, no. 6, pp. 2035–2040, Feb. 2009.

[36] S. C. Strother, Evaluating fMRI preprocessing pipelines, IEEE Eng. Med. Biol.Mag., vol. 25, no. 2, pp. 27–41, Mar.–Apr. 2006.

[37] N. Lange, S. C. Strother, J. R. Anderson, F. A. Nielsen, A. P. Holmes, T. Kolenda, R. Savoy, and L. K. Hansen, Plurality and resemblance in fMRI data analysis, Neuroimage, vol. 10, no. 3, part 1, pp. 282–303, Sept. 1999.

[38] N. Morch, L. K. Hansen, S. C. Strother, C. Svarer, D. A. Rottenberg, B. Lautrup, R. Savoy, and O. B. Paulson, Nonlinear versus linear models in functional neuroimaging:Learning curves and generalization crossover, in Information Processing in Medical Imaging (Lecture Notes in Computer Science), J. Duncan and I. Gindi, Eds.1997, pp. 259–270.

[39] A. S. Lukic, M. N. Wernick, and S. C. Strother, An evaluation of methods for detecting brain activations from functional neuroimages, Artif.Intell.Med., vol. 25, no. 1, pp. 69–88, May 2002.

[40] K. J. Worsley, J. Cao, T. Paus, M. Petrides, and A. C. Evans, Applications of random field theory to functional connectivity, Hum.Brain Mapp., vol. 6, no. 5–6, pp. 364–367, 1998.

[41] T. P. Minka, Automatic choice of dimensionality for PCA, Cambridge, MA:MIT, Rep.514, 2004.

[42] C. F. Beckmann and S. M. Smith, Probabilistic independent component analysis for functional magnetic resonance imaging, IEEE Trans.Med Imaging, vol. 23, no. 2, pp. 137–152, Feb. 2004.

[43] J. Mourao-Miranda, A. L. Bokde, C. Born, H. Hampel, and M. Stetter, Classifying brain states and determining the discriminating activation patterns:Support vector machine on functional MRI data, Neuroimage, vol. 28, no. 4, pp. 980–995, Dec. 2005.

[44] S. LaConte, S. Strother, V. Cherkassky, J. Anderson, and X. Hu, Support vector machines for temporal classification of block design fMRI data, Neuroimage, vol. 26, no. 2, pp. 317–329, June 2005.

[45] A. S. Lukic, M. N. Wernick, D. G. Tzikas, X. Chen, A. Likas, N. P. Galatsanos, Y. Yang, F. Zhao, and S. C. Strother, Bayesian kernel methods for analysis of functional neuroimages, IEEE Trans.Med Imaging, vol. 26, no. 12, pp. 1613–1624, Dec. 2007.

[46] D. R. Hardoon, J. Mourao-Miranda, M. Brammer, and J. Shawe-Taylor, Unsupervised analysis of fMRI data using kernel canonical correlation, Neuroimage, vol. 37, no. 4, pp. 1250–1259, Oct. 2007.

[47] S. C. Strother, S. La Conte, L. Kai Hansen, J. Anderson, J. Zhang, S. Pulapura, and D. Rottenberg, Optimizing the fMRI data-processing pipeline using prediction and reproducibility performance metrics:I. A preliminary group analysis, Neuroimage, vol. 23 (Suppl. 1), pp. S196–S207, 2004.

[48] J. Zhang, J. R. Anderson, L. Liang, S. K. Pulapura, L. Gatewood, D. A. Rottenberg, and S. C. Strother, Evaluation and optimization of fMRI single-subject processing pipelines with NPAIRS and second-level CVA, Magn.Reson.Imaging, vol. 27, no. 2, pp. 264–278, Feb. 2009.

[49] J. Zhang, L. Liang, J. R. Anderson, L. Gatewood, D. A. Rottenberg, and S. C. Strother, Evaluation and comparison of GLM- and CVA-based fMRI processing pipelines with Java-based fMRI processing pipeline evaluation system, Neuroimage, vol. 41, pp. 1242–1252, July 2008.

[50] R. Kustra and S. C. Strother, Penalized discriminant analysis of [15O]-water PET brain images with prediction error selection of smoothness and regularization hyperparameters, IEEE Trans.Med.Imaging, vol. 20, no. 5, pp. 376–387, May 2001. [SP]

Ron Schneiderman

DSP 在消费电子应用的演化

您将很难找到一个不需要数字信号处理消费类电子产品。消费电子是一个很大的市场　今年市场总值大概有 1650 亿美元(参见　消费电子市场增长和创新的形势如何？），随着新的创新产品的推出，DSP 的需求也持续增长。

随着纠错码这一发明的出现，数字信号的价值在 1948 年日益显现，纠错码不仅能传输信号，而且在传输过程中检测并纠正错误。

就在同一年，贝尔实验室宣布发明了晶体管。也就是在这一年，哥伦比亚广播公司实验室的负责人 Peter Goldmark 博士，因为不得不在他最喜欢的古典音乐作品片段中间翻转这种 78 转的唱片而恼火，就发明了密纹（LP）慢转唱片。Ampex 也开始在 1984 年售卖磁带录音机（当然是盘式的）。

但是，DSP 真正开始成为消费电子产品的重要组件是在 20 世纪 70 年代。

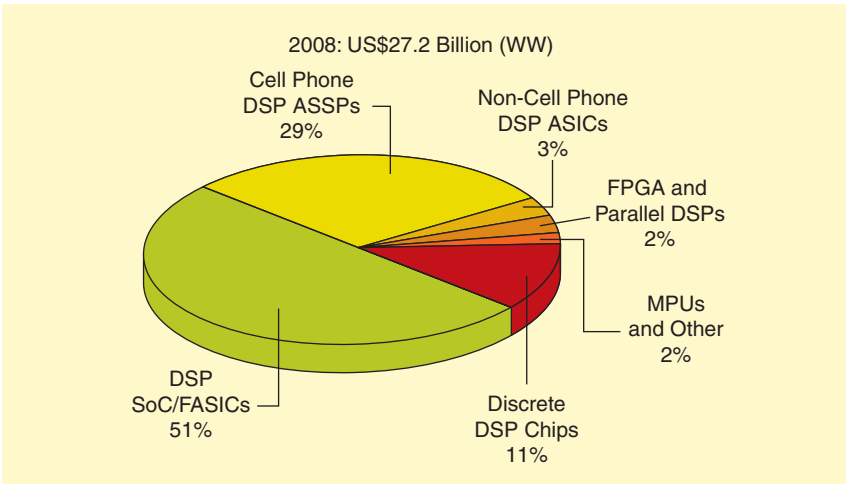
第一台个人电脑 Altair 8800，于 1975 年开始以工具包的形式出售，之后在1997年出现了组装电脑，苹果II。Sony 和 JVC 也在 1975 年开始销售录像机（JVC 采用 VHS 格式，而 Sony 采用 Betamax 格式）。随着德州仪器（TI）在 1978 年推出的一款玩具，DSP 进一步地渗透进消费市场的。这款玩具名为 Speak & Spell，玩具的特点是采用了一个 DSP 特定语音合成芯片，通过单词的发音教孩子拼写，并指出他们的拼写是否正确。还有其他的公司也在设计和生产单片的 DSP。（Intel 在 1979 年推出了一款单片 DSP。）

Philips Electronics 因其在光盘（CD）发展方面所做的贡献，获得了 IEEE 里程碑奖首先是在 1979 年出现的一个代号为 Pinkeltje 的原型装置，这是在首个进入千万个消费者家庭的大规模数字化消费产品，也是受益于数字信号技术的首个消费产品。

> 随着新的创新产品的推出，DSP 的需求也持续增长。

到了1985年，根据 Fifty Years of Signal Processing：The IEEE Signal Processing Society and Its Technologies 1948–1998 的说法，对于 DSP 芯片而言，只有三个大的商业市场　语音编码、视频压缩和调制解调器。总共会有价值 5000 万美元的市场。自那时起，DSP 进入了电子行业的角角落落，新的和创新的消费电子产品和应用的发展带来了极大的增长（图1）。

根据专门跟踪和分析数字信号处理市场的 Forward Concepts 的说法，2008 年消费品部分占了数字信号处理市场 65 亿美元 中的 8.37 亿美元。Forward Concepts 预测全球数字信号处理芯片市场在 2008 年到 2013 年这五年间，将以 9.4% 的速度增长，达到 430 亿美元的水平；但是，从 2009 年开始，四年的增长率预测为12.1%（年复合平均增长率）。



[图1] 随着目前许多 DSP 芯片被称为或标识为 SoC 类，例如 ASIC 和 ASSP，有关 DSP 的消费电子产品市场正在发生改变。现成的或“分立 DSP 芯片”只占 DSP 芯片世界的一小部分。事实上，分立 DSP 现在大约只占 270 亿美元市场的11%。（引用得到了 Forward Concepts 公司的许可。）

数字对象标识符 10.1109/MSP.2010.936031

但是 Forward Concepts 的创始人和总裁 Will Strauss 说，无线和消费电子产品将以更快的速度增长（图2）。

市场、技术的挑战

但是，市场和技术都是在不断变化中。

起初，许多组件都曾被称为 DSP 芯片 ，现在更多地是被贴上了片上系统(SoCs)的标签，就像专用集成电路（ASIC）或专用标准产品(ASICs)一样 即使是传统的数字信号处理芯片供应商也这么做。因此，今天所谓的离散数字信号处理芯片是数字信号处理芯片市场的一小部分，几乎没有被算作 DSP 芯算。不过，Strauss 说，数字信号处理作为推动整个半导体市场的一项技术，正在为应用设计和开发工具创造一个成长中的市场。

随着今天高度集成的芯片设计和对这些先进处理器的编程越来越复杂，开发工具的选择是目前 DSP 选择的关键 Strauss 说。

硅知识产权 DSP 内核和消费电子平台解决方案的领先授权厂商 CEVA 随后采纳了 Strauss 的评析，整合优化工具链，启用了终端到终端，完全基于 C 的可授权 DSP 内核开发流程。CEVA 声称开发将大大提高产品的整体性能，并为缩短 SoC 设计的周期。

据报道，许多公司使用专有 DSP 内核都转而使用 CEVA 内核，特别是手机应用程序。这种趋势的原因是芯片价格的压力，已促使芯片设计公司减少或卖掉其手机芯片生产线，同时加大来自基带芯片供应商的竞争。

DSP 在消费类电子产品中的蓬勃发展

尽管移动电话代表了数字信号处理器芯片（如基带和应用处理器）最大的单一市场，这也是音频设备的重要组成部分。

消费电子市场的发展和创新是什么样的？

就各方面而言，消费电子都是电子工业的不可小觑的一部分。

根据由美国消费电子协会（CEA）每半年一次的行业预测，消费电子产品今年在美国产生了超过 1650 亿美元的年销售收入，比 2009 年（该年份的行业收入在20年来首次下降）略有增加。

移动手机有望成为该行业的主要驱动力。 智能手机占手机总出货量超过30%，在2010年创造了近 170亿 美元的销售收入，销量超过 5200 万台，预计这一数字在今后几年还会增加。（诺基亚表示，2010年1月份预计销量超过 500 万台，远高于市场预期，预计占有对全球40%的市场份额。）

在2010年，电脑的销售预计也会不错。上网本的销量比 2009 年翻了一番，这个相对较新的电脑类别与以前的预测相比，销售更加强劲。CEA 的预测 2010 年将会售出超过 3000 万台上网本，产生超过 140 亿美元的收入。CEA 的行业分析总监，Steven Koenig 说： 智能手机和上网本具有强劲的增长潜力，原因是消费者在孜孜不倦寻求高效、便携的设备 。

过去数年来，随着消费者向高清平板电视过渡，电视机市场也一直都是收入的主要驱动力之一，CEA 认为今年的销量将超过 3700 万台。创新的电视显示器，例如三维（3D）、互联网连通、以及有机发光二极管技术，有望继续保持增长，而且有助于维持显示器类的收入。CEA 预计在 2010 年 3D 电视的销量超过 400 万台。

对于消费电子厂商而言，汽车行业也将继续发挥更大的作用，汽车制造商使用电子设备，让他们的产品与众不同。

创新在消费电子产品的成长历程中举足轻重。一月初，在拉斯维加斯举行的国际消费电子展（CES）上，有超过 2500 家科技公司引进了超过 20,000 个新产品，其中 CEA 公布了一项 Zogby 调查结果，调差显示，96% 的美国人认为创新是美国努力在全球经济舞台上保持领舞的关键。

CEA 的总裁和 CEO，Gary Shapiro 先生说 我不知道还有其他别的事情能得到 96% 的美国人的赞同。

作为 Qualcomm 的主席和 CEO，Paul Jacobs 博士在 CES 期间所做的主题演讲中提到， 无线与消费电子的契合将会以出人意料的方式进行 ，因为越来越多的消费电子设备很快就具备了手机的功能。

期待新的无线产品和应用出现，可能会有助于在大范围对无线运营商的业务进行预期的重新分配。

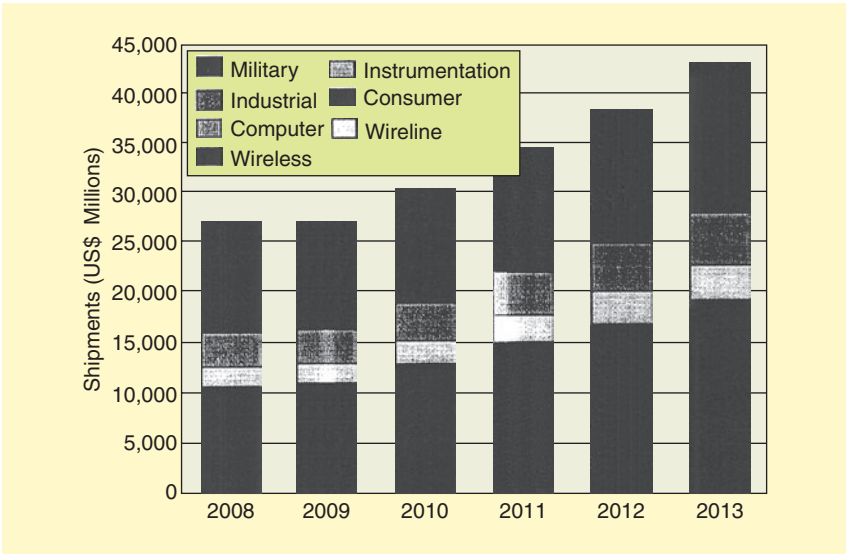
虽然无线设备的增长是全球性的，但大多数分析家预计北美市场的增长会格外强劲，北美的消费者已经习惯了拥有多个无线设备。

有几家公司活跃在这个市场。

Cirrus Logic 推出了一款 SoC，CS47048，主要针对音频放大器应用。

将一个 32 位的音频 DSP，高性能多声道音频编解码器和一个数字音频接收/发射器集成到一个芯片（IC）上，大幅降低了整体电路板空间要求和系统成本。

Cirrus 目前正在开发一个新的音频 SoCs 产品线，结合了音频 DSP 和不同级别的混合信号内容，并使用的 S/PDIF 接收器和



[图2] 无线和消费电子产品是全球范围内 DSP 出货量增长最快的两大块市场。（引用得到了 Forward Concepts 公司的许可。）

S/PDIF 发射器。Cirrus 音频 DSP 产品大多针对消费应用，重点在于音频，而不是通用数字信号处理器。

虽然 National Semiconductor 不生产 DSP 芯片，但将数字信号处理器集成到其移动电话的混合信号音频系统。

National 最近扩展了两个新器件的低功耗 Boomer D类音频子系统产品系列，旨在简化便携式产品的设计。LM49352 内置音频编码解码器、接地参考的头戴式耳机放大器、小型听筒驱动器、D类扬声器及音频数字信号处理器。

National 还提供结合远场噪声抑制技术，降低背景噪音，改善移动电话和供电耳机中语音通讯清晰度的音频产品线。National 声称 PowerWise 产品线只消耗具有可比性的数字信号处理器软件系统的功率（1 mA）的十分之一，无需增加数字信号处理器或微处理器的语音处理程序代码的编写和测试开发时间。

Tensilica, Inc 正致力于 MIPS 技术，以推动 Google 的 Android 平台的 SoC 设计。Tensilica 和 MIPS 在一月份的国际消费电子展商共同展示了一个集成了 Tensilica 的 HiFi 2 音频 DSP 的处理器内核。Samsung 最近获得 Tensilica, Inc. 的 HiFi 2 音频 DSP 的授权，将其用于 Samsung 下一代多媒体系统产品。Tensilica 支持数据层面处理器产品，包括蓝光光盘播放机，蓝牙功能设备，液晶显示电视机，手机，WiFi 和无线通用串行总线功能的笔记本电脑，无线高清晰度多媒体界面，手持式游戏机，与喷墨和激光打印机。

设在日内瓦的意法半导体公司（STMicroelectronics）也在高清 SoC 芯片中使用专用的。双重可编程音频 DSP，投放全球平板电视市场。但在今年一月，STMicroelectronics 推出一款单核的多触点 电阻式 触摸屏控制器（作为其多触点产品 新 STMTouch 系列的第一个成员）和接近和触摸键感应器，STM 微控制器部的总经理 Jim Nichols 称其为一个具有附加价值的解决方案，与其他需要专门的编程知识的 昂贵的多核处理器或数字信号处理器 相比的话。Nichols 说，新的微控制器的开发是为了支持日益复杂的应用和智能手机游戏、移动互联网设备和上网本。

Nichols 的评论似乎在呼应去年德州仪器说法 对于许多设计师而言，评估一个新的 DSP 平台时，成本和时间，设置开发工具成为了一个主要障碍。TI 的回应是推出其 eZdsp USB 记忆棒开发工具，将全功能仿真器和集成开发平台的成本降至 49 美元。TI 表示，这会加速 DSP 应用的创建，包括便携式音频播放器、录音机、IP 电话、便携式医疗设备、生物识别 USB 密钥、软件定义的无线电设备、免提耳机及计量应用。eZdsp 无需其他组件或电缆，整个开发工具由 USB 端口驱动。设计人员只需将其插入任何笔记本电脑或工作站的 USB 端口即可。

今年1月，TI 还增添了两个新的器件，最低功耗的 16 位 DSP 平台，声称高度集成，可提升便携式设备性能的 20%。该战略旨在使客户能够维持非常低的功率水平，同时还增加了诸如额外的语音，音频的编解码算法，以及便携式通信和应用的功能。典型的转变已不仅仅考虑独立的数字信号处理器，该器件还集成了电源管理功能，如一个片上低压降稳压器，以及动态电压和频率缩放，以使设计人员能够最大限度地发挥和有效管理电池寿命便携设备。（TI 还继续大力推广其高端三核数字信号处理器，主要用于无线长期进化的原始设备制造商（OEM）网络中使用的蜂窝通信基础设施。）

同样，Marvell 半导体最近推出了针对使用 ARM 指令集的消费电子产品的应用的 四倍 核处理器。Weili Dai，Marvell的创始人之一，副总裁以及该公司的消费品和计算业务部总经理说，四核实现的话，每核可以提供超过千兆赫的处理能力，是专门针对诸如大众消费市场和高容量游戏应用的客户专用产品。

另一个在音频领域，以数字信号处理为基础的项目就是 GN Netcom 公司的 Jabra Cruiser，用

于手机的蓝牙无线扬声器，具有噪音阻断技术，以及消除交通噪音的双麦克风系统。

Jabra 技术使用双麦克风捕捉声音，然后只过滤掉环境噪音。和 DSP 一起使用，用于降低回音，该技术允许通话双方都能以接近自然的语音质量听到声音。

十大消费移动应用

随着移动手机引领 DSP 市场增长方式，哪些应用（应用程序）将在移动市场增长最快？

作为领先的信息技术研究和咨询机构，Gartner 公司制作了它认为是 2012 年十大消费移动应用的名单。这份名单是根据其对消费者和业界的影响，考虑收益、消费忠诚度，、商业模式、消费者价值、估算的市场占有率而得。

根据 Gartner 的说法，2012 年十大消费移动应用包括：

1)移动转账业务：这项业务使得人们能够汇款给使用短消息服务（SMS）的其他人。其低成本，更快捷和便利的特点对发展中市场的用户很有号召力。但这一业务也面临挑战，包括监管的风险。

2)定位服务(LBS)：LBS 是上下文感知服务的一部分，Gartner 公司预期其在接下来的数年内会成为最具颠覆性的技术之一。根据 Gartner 的估计，2009年全球LBS用户将超过9600万，2012年将达到5.26亿。

3)移动搜索:Gartner 公司表示，移动搜索列在十大业务的第三位，是因其对技术创新和行业收入有很大的影响力。Gartner 预计移动搜索的忠诚度将在若干移动搜索运营商间分摊，这些移动搜索提供商在技术上会有其独特之处。

4)移动浏览：移动浏览位列第四原因是它在商业领域的广泛应用。2009年，在全球出货的手机中，60% 具有移动浏览功能。根据Gartner的预测，到2013年，这个比例会上升到80%。

5)移动健康监控：今天，移动健康监控还处于初级阶段，发展也很缓慢，但可以看到有非常大的潜力，因为移动性的移动网络覆盖比固定网更重要，尤其是在发展中国家，。

6)移动支付：在可用的支付方式很少时，移动支付可以作为一种付款方式。这是在线支付的一种扩展，也是加强安全认证的附加要素。主要源于对多方面业务的影响，包括银行、零售商、消费者以及移动运营商。缺点是不同的技术和商业模式的实现可能会创建一个非常分散的市场。

> 创新在消费电子产品的成长历程中举足轻重。

7)近场通信(NFC)服务：近场通信可实现相互兼容装置间的无线数据转输，只需将它们放在靠近的地方（10CM）。这一技术可用于零售购买、交通、个人识别和信用卡。NFC 最大挑战是达成移动运营商和服务供应商的商业协议。Gartner 预计从 2010 年下半年开始 NFC 会有大规模的部署。

8)移动广告：被看做移动互联网上内容货币化的重要途径，尽管经济衰退，手机广告业务还是继续增长。2008年，移动广告总支出是5.302亿美元，2012年，这个数字可以达到 75亿美元。

9)移动即时信息(IM)：价格和可用性问题已经阻碍了移动即时通讯的广泛应用，但 Gartner 认为用户需求和市场条件，将引导未来移动即时信息的发展。移动 IM 被认为是移动广告和社交网络发展的一个机遇，已经内置于某些较先进的移动 IM 客户端中。

10)移动音乐:：虽然至今市场部分还是有些令人失望（除了手机铃声和回铃外，这是个可产生数以百万计收入的业务），消费者表示他们希望他们的移动电话有音乐相伴。新的创新模式和服务计划有望在 2012 年成为一个增长点。

消费者的移动应用和服务不再是移动运营商的特权，Gartner 的研究主管，Sandy Shen 如是说。 越来越多消费者青睐智能手机，互联网企业都参与到移动业务中来，应用程序商店存储和跨产业服务的出现降低了移动运营商的主导地位。每个企业都会对消费者如何交付和体验应用程序产生影响，那就看谁能最终获得他们的注意力和消费力。

这些应用中的某些或全部都可能对 DSP 市场产生影响，如何发挥技术的影响，以及如何将其整合到移动产品的新的，不断增长的范围内。 [SP]

Steve Young

认知型用户界面

智能人机交互–有应用程序可行吗？

本文提出的观点是：未来一代计算机系统将需要使用 认知型用户界面 来实现足够鲁棒和智能的人机交互。这种认知型用户界面的特点是，具备推理能力、能在具有不确定性的情况下进行策略规划，在短期可以进行自适应调节，在长期可以根据经验进行学习。局部可观测马尔科夫决策过程(Partial Observable Markov Decision Processes: POMDP) 是实现这种界面的一种合适的工程框架。这种框架结合了贝叶斯置信跟踪技术和基于收益值的强化学习技术。它的好处可以通过后文一个简单的采用触摸手势驱动的 iPhone 应用程序界面的例子来说明。而且，证据表明，人类似乎对于不确定条件下的规划也使用相似的机制。

POMDP 框架的一个局限性是难以处理精确计算，因此，POMDP 往往被认为对实际问题是不切实际的。本文的第二部分将说明POMDP 方法最核心的优势可以通过使用合理的近似算法而在解决实际问题的过程中得以保留。为了说明这一点，我会详细讨论两个用于实际情况的口语对话系统(SDS)。每个系统都有很不同的近似算法，但是都能实现显著的性能提高。第一个称为 隐信息状态 (Hidden Information State: HIS) 系统，它是传统口语对话系统的一种自然扩展。第二个是： 对话状态的贝叶斯更新 (Bayesian Update Dialogue State: BUDS) 系统，它采用了贝叶斯网络理论研究中的一些最新成果，虽然在系统规模的扩展性方面仍然面临问题，但它却为短期和长期的系统自适应提供了更多的可能性。文章最后指出，尽管认知型用户界面的未来发展所面临的挑战是巨大的，但使用这样的界面是必然的趋势。

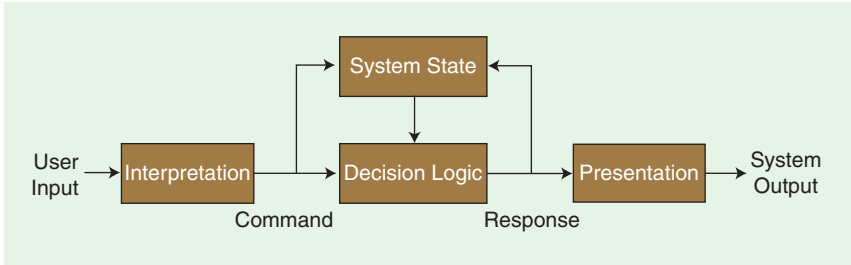
PHOTO BY MATTHIEU BARRAGUÉ

引言

随着计算机系统的复杂性不断增加，对更加鲁棒

数字对象标识符 10.1109/MSP.2010.935874

和智能的人机交互的需求将日益增长，并将最终超出目前通用的传统人机界面技术可以支持的范围。而且，这种推动技术发展的需求事实上已经出现。高性能触摸屏智能手机的引入催生了非常先进的新一代手机应用模式，即用户必须通过使用触摸手势和语音的组合完成复杂的交互。然而，手机屏幕通常较小，且环境噪音通常较高。因此，保持一个可接受的鲁棒性水平将是一个重大的挑战。为大众市场开发智能机器人，例如为老年人提供帮助等，将进一步提出新的挑战。在这里，语音控制将是至关重要的。但一系列的技术问题会不可避免的出现，如信号源分离的不充分、不可靠的语音端点检测、语义理解错误、用户自身意图的不确定性和对计算机系统的理解混淆等。这些因素的存在都为对用户意图的准确可靠的解读增添了困难。另一个例子是迅速扩展的电脑游戏业（在美国电脑游戏业的营业额现在已经超过了电影业）。有关人类玩家与计算机生成的人物进行真实对话的全新一代 身临其境 游戏，将促使我们发挥最大的能力去创建鲁棒和自然的用户界面。虽然与游戏相比虽然不是那么重要，但在医疗保健支持和教育领域，对话系统方面也会出现同样具有挑战性的应用实例。



[图1] 有限状态的人机交互自动机模型。每个用户输入都被视为一个命令，根据一定的决策逻辑，这个命令使自动机从一个状态转移到另一个状态。每次进入一个新状态的时候，自动机会对用户产生一个响应。.

这篇文章是在 2009 年 IEEE 声学、语音学和信号处理国际会议上的完整讲稿。它的基本前提是：未来的人机界面必须满足以下四个关键特征才能应对上述挑战：

*1) 支持推理和推论的能力。*自然的人际交往通常依赖于不精确的模拟信号，诸如手势、面部表情和语音。人机界面必须有能力根据上下文解释这些输入，稳健地解决这些含糊之处，尽量减少失误。

*2) 在不确定性条件下进行规划的能力。*有效的沟通往往可以利用不完整的知识实现特定的沟通目标。这就要求定义客观的沟通目标，并通过对话交互策略的优化，尽可能有效的满足目标。

*3) 在线适应变化的能力。*对话环境是不断变化的，人机界面必须能够即时改变自身的运行方式，以维持可接受的性能水平。

*4) 从经验中学习的能力。*除了短期内适应，从长期来看，人机界面应该能够从它自身与用户的互动中学习更具一般性的知识和行为方式。越是使用，它应该变得越聪明。

具有上述四个基本属性的人机用户界面将被称为 认知型用户界面 。

几乎目前所有的人机界面都采用图1所示的有限状态自动机模型。所有相关信息都被编码在一个有限状态机内。每个用户输入都被视为一个命令，根据一定的决策逻辑，这个命令使自动机从一个状态转移到另一个状态。每次进入一个新状态的时候，自动机会对用户产生一个响应。把这个模型应用于洗衣机的简单按钮界面或者基于语音的信息查询系统的复杂自然语言界面是等价的。唯一的区别是，后者中存在的模糊性要高得多，因为语音输入信号往往无法正确识别。然而，两者的操作都基于状态完全已知的假设。

事实上，这个假设永远不会被满足。即使在一台洗衣机上按下一个按钮可以是完全明确的，但却可能无法代表用户的真正意图。人类经常在不确切的意图下，基于不完全的信息来进行交流。因此，用户意图中总存在着某种不确定性，而不确定性的问题会由于与人类互动的 IT 系统过于复杂而变得更为严重。对于这样的交互系统，越来越多的引入不精确的多模式输入（例如手势、情感特征、目视特征和语音），以及鲁棒的处理不确定性的机制就变成了一种不可阻挡的趋势。

这篇文章中提出的主张是明确而直接的：不确定性无法避免，未来的交互界面如果要用于某个具体的目的，则必须具有认知能力。接下来的部分将论证基于贝叶斯推理和贝尔曼最优性准则的 POMDP 框架是建立下一代认知型用户界面的一种适当的工程方法。

应该指出的是，这是一篇表明立场的文章，而非评论性文章。因此，文章中适当地给出了参考文献，但是没有全部给出。

一个例子：简单的触摸手势驱动界面

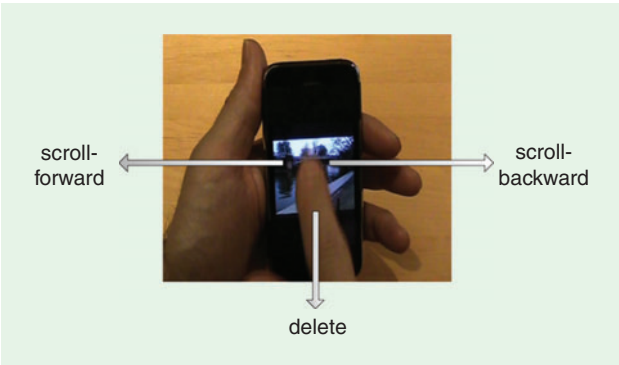
iPhone 是一个很好的例子，可以显示直观的触摸手势驱动界面是如何提高我们与设备间沟通能力的。不过，有些操作并不如我们希望的那样灵活。假设一种情况：您已经拍摄了大量的照片，想快速浏览它们并删除其中一些照片。默认的界面要求您

选择每张照片，按下删除按钮，然后明确地确认每个操作。如图 2 所示，做这件事情的一个快捷方式可能只使用 3 个手势，即前滚、后滚和删除。这种界面的唯一问题是，当您尝试快速进行时，您的手势变得不可靠，从而产生错误。当然，在实际系统中，将需要某种形式的恢复机制以防止意外删除，但是，这里所关注的是从一开始就要尽量减少出现这样的错误。

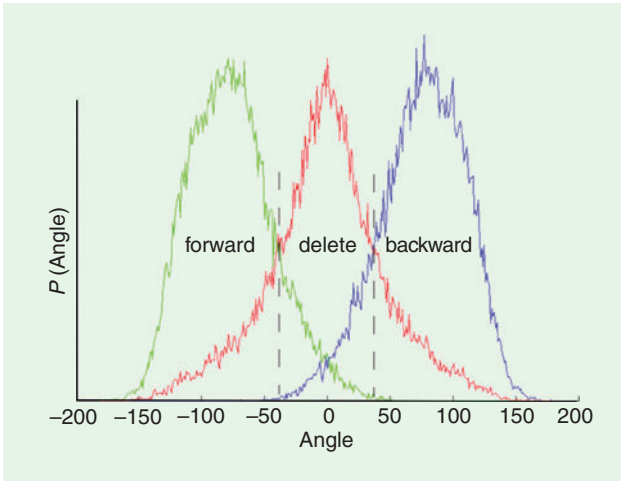
> 切合实际的近似，可以保留POMDP方法的根本优势.

首先考虑用经典的方法来实现图1 的交互界面。假设每个手势可以通过屏幕上手划曲线的角度来识别（如图3所示）。最经典的方法需要两个阶段：首先是将每个手势的角度识别为上述三个可能的命令之一；其次是将识别结果输入到决定系统响应的某种决策逻辑中去。

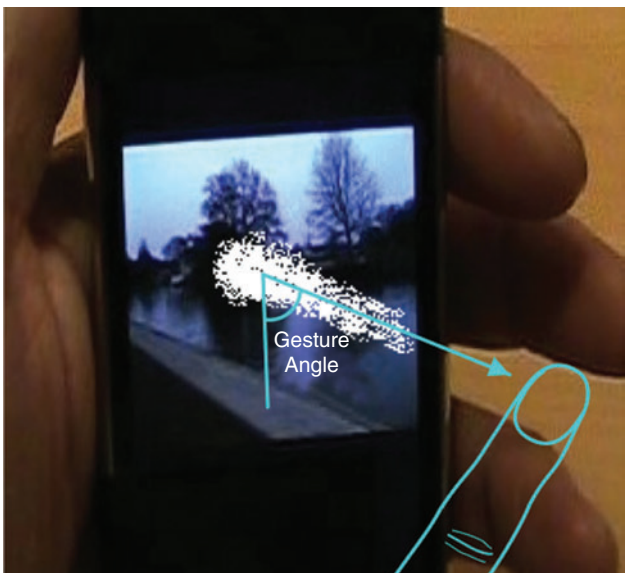
这些步骤的首要问题是一个分类问题。假定手势与竖直方向的夹角为θ，由于可能存在检测错误，通常我们会对每一类命令$\omega = (forward, delete, backward)$（见图4）估计一个概率分布$P(\theta|\omega)$。最优决策的边界$\theta_i^*$可以根据类后验概率$P(\theta|\omega)$来确定。图4给出了在这种情况下是一个典型分布，这里的平均误$\theta \approx 20\%$。选择的一种典型方法是要求在决策边界上的后验概率相等，如图4中的垂直虚线。对于每一个输入手势，识别出其角度之后，可以通过与决策逻辑中的角度阈值进行比较来选择适当的系统响应操作。作为进一步的完善，决策错误率的概率分布可以通过角度的后验分布估计出来，并同时估计一个置信边界δ。因此。因此，经典方法的第二阶段通常以一个简单的程序或流程图的形式概述，如图5所示。



[图2] iPhone照片选择使用的三个手势：a）摁左键向前翻阅照片；b）摁右键向后翻阅照片；c）摁向下键删除照片。.



[图4] 每个命令的手势角度的经验分布。给定一个手势的角度，通过将其与如虚线所示的类边界比对，确定最有可能的命令。分布之间的重叠程度决定了错误率。



[图3] 识别一个手势。每个手势通过它在屏幕上的角度来识别。

```
 1: Let θ = angle of input gesture and
    θ1 and θ2 be lower and upper thresholds
 2: Let δ be a confidence margin around
    each threshold
 3: if θ < θ1 - δ then
 4:     scroll-forward
 5: else if θ1 + δ < θ < θ2 - δ then
 6:     delete-photo
 7: else if θ2 + δ < θ then
 8:     scroll-backward
 9: else
10:     do-nothing
11: end if
```

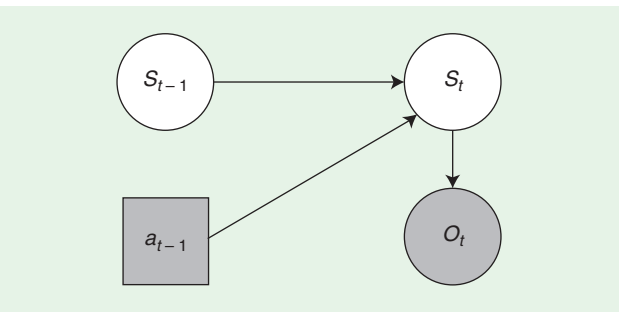
[图5] 识别每一个手势的决策逻辑。通过引入一个以决策边界为中心，宽度为2δ的边距来降低误差的影响。当一个手势落在这个边距内，该命令将被忽略。

那么，当用这种方法去设计用户界面时，丢失了什么？首先，没有明确的描述不确定性的模型。如上例所示，置信边界虽然可以用来辅助决策，但是，角度识别过程本身仍然是输出一个确定性的结果，而且一旦这个结果被后面的决策过程采用，就无法轻易去除影响。第二，没有尝试跟踪用户的意图。因此，该系统无法确定它对用户手势的解释与用户要做什么是否一致。在这个例子中，系统可能会观察到用户很少在删除照片后还往回浏览，但是，在往回浏览时，最有可能的下一个手势就是删除照片。这种行为特征完全可以用来消除手势角度识别不准确产生的歧义。第三，由于没有量化指标，不可能对流程图中的决策规则进行优化。这一切的后果是不能满足认知型用户界面所需的标准。

建立一个认知型用户界面的关键是认识到在解释手势输入时会出现不确定性，因此，不将它们视作确定的命令，而把它们视为一种观察到的特征，利用这些观察特征，系统可以推断出用户的意图。系统响应用户意图的有效性可以通过一组收益值来量化，必要的决策逻辑可以通过最大化这些收益值而达到最优。这种方法的工程实现依赖于两个基本思想：贝叶斯推理和贝尔曼最优化原则，这个框架就是通常所说的 POMDP [2], [3]。

回到iPhone的例子，在每一个时刻，用户有三种可能的意图：前滚、后滚和删除照片。这些意图由一个离散的状态来表示，即 $s=\{$向前、删除、向后$\}$。在此，应注意这些状态是指用户的实际意图状态，而不是计算机系统的状态。为满足用户需求，机器提供四个可能的操作：$a=\{$前滚、删除照片、后滚、什么也不做$\}$。

用户 t 时刻的意图 s_t 取决于其先前的意图 s_{t-1} 和当前系统的操作 a_{t-1}。因此，用户的意图变化可以通过转移概率 $P(s_t \mid s_{t-1}, a_{t-1})$ 获得。t 时刻产生的手势特征 o_t 将仅取决于用户当时的状态 s_t。因此，可以通过概率 $p(o_t \mid s_t)$ 密度分布来描述用户表达特定意图时采用的各种可能方式。请注意，观察特征仅仅是检测到的简单的手势角度特征，在这里，我们并不像前面经典方法中那样要对手势特征进行分类识别。



[图6] 以手势界面为例，给出了一个时刻的贝叶斯网络。隐藏的系统状态s由圆圈表示。由阴影表示观察o和操作a。

当然，这里的关键问题还是在于用户的实际意图是无法直接观察的，它是一个隐变量，其取值只能从状态转移概率，特定状态下的观察特征概率分布函数，以及实际观察到的角度特征来推断。这些关系可以通过一个如图6所示的贝叶斯网络来描述。图6中，圆圈代表隐变量，用阴影表示的圆圈代表可观察的变量，方框表示机器的操作[4]。

令 $t-1$ 时刻的隐状态 s_{t-1} 的特征分布为 $b_{t-1}(s_{t-1})$，则所谓 推理 就是要在给定 b_{t-1}，a_{t-1} 和 o_{t-1} 的条件下寻找 $b_t(s_t)$。这一问题很容易通过贝叶斯公式解决

$$
\begin{aligned}
b_t(s_t) &= P(s_t \mid o_t, a_{t-1}, b_{t-1}) \\
&= p(o_t \mid s_t) P(s_t \mid a_{t-1}, b_{t-1}) / p(o_t \mid a_{t-1}, b_{t-1}) \\
&= p(o_t \mid s_t) \sum_{s_{t-1}} P(s_t, s_{t-1} \mid a_{t-1}, b_{t-1}) / p(o_t \mid a_{t-1}, b_{t-1}) \\
&= k \cdot p(o_t \mid s_t) \sum_{s_{t-1}} P(s_t \mid s_{t-1}, a_{t-1}) b_{t-1} p(s_{t-1}),
\end{aligned}
\tag{1}
$$

其中，$k = 1/p(o_t \mid a_{t-1}, b_{t-1})$ 是归一化常数，而状态相关的特征分布函数往往是由一个称作 置信状态 的 N 维向量 $b = [b(s_1), \cdots, b(s_N)]$ 来表示。于是，置信状态的更新就可以写成如下矩阵形式：

$$
b_t = k \cdot O(o_t) T(a_{t-1}) b_{t-1} \tag{2}
$$

其中，$T(a)$ 是对于系统操作 a 的 $N \times N$ 转移矩阵，$O(o) = diag([p(o \mid s_1), \cdots, p(o \mid s_N)])$ 是特征概率分布的对角矩阵。因此，进行一轮推理（包括进行归一化）的时间复杂度是：$O(N^2 + 3N)$。以简单的 iPhone 为例，其中，$N = 3$ 是完全可控的。然而对于更复杂的情况，由于 N 很大，将难以做出准确的计算。这一主题将稍后作出详细讨论。

给定一组初始值 b_0，经过对每个手势的连续观察，通过(2)式，置信状态就可以不断更新。因为实际的确切状态是不可知的，在每一回合对话中，系统所采取的操作就必然基于置信状态而不是那个未知的隐状态。从置信状态到操作的映射取决于策略 $\pi = b \to a$。任何特定策略的好坏都可以通过指定所有可能的 状态-操作 组合的收益值 $r(s,a)$ 来量化。以iPhone为例，表1给出了可能的收益值。表1

[表1]对每个可能的状态-操作对的收益值

		操作			
		后滚	删除照片	前滚	什么也不做
状态	向后	11	220	21	0
	删除	21	15	21	0
	向前	21	220	11	0

[表2] 转移矩阵$P(s'|s,a)$。对于一个特定的动作，每一个3 3的网格对应一个状态转移矩阵。在列标签b、d和分别代表了状态后退、删除和前进。

		STATE s'											
		b	d	f	b	d	f	b	d	f	b	d	f
STATE s	backward	1	0	0	0.3	0.4	0.3	1	0	0	1	0	0
	delete	1	0	0	0	0	1	0.1	0.4	0.5	0	1	0
	forward	0.1	0.4	0.5	0	0	1	0.2	0.3	0.5	0	0	1
ACTION a		scroll-forward			scroll-backward			delete-photo			do-nothing		

中，对符合用户意图的操作给予了积极的回馈，对不符合用户意图的操作给予了相应的惩罚。

从用户的角度来看，不正确的删除应给予最强的惩罚，因为意外删除照片是系统犯的最糟糕的错误。

收益值的具体形式的选择是一个设计决策，不同的收益值会导致不同的策略和不同的用户体验。收益值函数的选择可能也会在策略优化期间影响学习率。但是，一旦收益函数固定，策略的质量则要通过用户交互过程中总收益的数学期望来衡量，即策略优化等价于最大化R。

$$R = \varepsilon\{\sum_{t=1}^{T}\sum_{s} b_t(s) r(s,a_t)\} = \varepsilon\{\sum_{t=1}^{T} r(b_t,a_t)\} \quad (3)$$

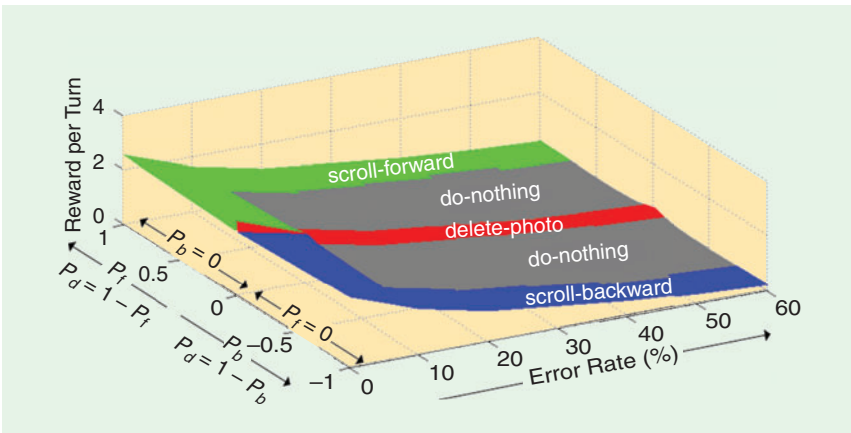
如果整个过程具有马尔科夫性，则采用策略 从任意一个给定的置信状态b到到交互的结束状态的总收益 将独立于其前面的所有状态。使用贝尔曼的最优性原则，可以通过迭代计算出这个价值函数的最优值。

$$V^*(b) = \max_a\{r(b,a) + \sum_o p(o|b,a)V^*(\tau(b,a,o))\} \quad (4)$$

其中，$\tau(b,a,o)$表示式(2)中定义的状态更新函数[5]。这种迭代优化是强化学习的一个特例[6]。

这种最优值函数对有限的交互序列而言是分段线性且具有凸性的。它可以表示为n维超平面张成的置信空间中的一个有限集合，其中，集合中的每个超平面对应一个相关的操作。这种超平面的集合还定义了最优策略，因为对于任何置信状态b，我们所需做的就是要找到具有最大期望值$V^*(b)$的超平面，然后，选择对应的操作[3]。

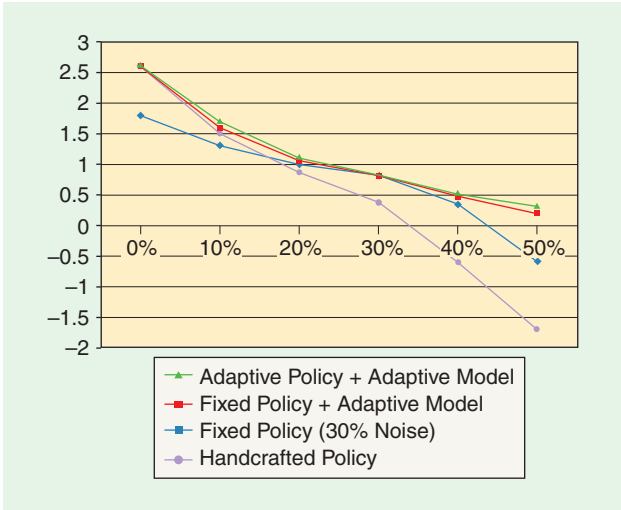
可以从图4所示分布中估计观测概率矩阵O。注意，O将取决于用户手势的识别错误率。为了反应这个情况，在例子中，观察矩阵的 7 个离散错误率的范围是从0%到60%。在此，如果一个手势的角度位于最低错误决策边界的错误的一边，即$p(W_{intended}|\theta) < p(W_{not-intended}|\theta)$，我们即认为该手势有误。给定$T$值，$O$值和收益值函数，可以使用贝尔曼最优性原则来优化策略。如前所述，策略是一组超平面集合，这些超平面的上表面定义了最优值函数。随着手势的错误率的增加，这个上表面的复杂度也会增高。对于这里的例子，在0%错误率时，策略仅包括三个超平面，而在60%错误率时，策略包括约37,000个超平面。图7总结了在七种不同的错误率下学到的策略。在这里，假定当$P_f = P(forward)$非常大时，$P_b = P(backward) = 0$，这样置信空间就被压缩成一个一维空间。因此，当在 向前 和 删除 状态中进行选择时，图的后部显示出价值函数的表面；而当在 删除 和 向后 状态中选择时，前部显示价值函数的表面。表面的颜色表示了在置信空间中的任何点采取的最优系统操作。可以看出，在 0% 的错误率时，除非删除的可能性非常接近 1，否则决策将会选择向前或向后滚动。在高错误率时，一个什么也不做的操作区域被引入，以避免无意中删除照片。这也表明，当错误率上升时，价值函数本身（表示为每轮的平均收益值）会稳定下降。



[图7] 对于不同的手势错误率，在一个压缩置信空间上画出了策略价值的函数。横轴一分为二，表示了用户希望前滚的概率(Pf)、希望后滚的概率(Pb)和希望删除的概率(Pd)。 在左半部分，Pb=0，当Pd从0增加到1时，Pf从1减小到0。右半轴是其镜像，即Pf=0，当Pd从1减小到0时，Pb从0增加到1。其他横向维度指错误率，垂直轴是每轮平均的收益值。表面的着色是指沿着置信状态维在每一点采取的最佳操作。

iPhone 照片排序程序的性能可以通过对状态转移和观察概率模型进行采样以模拟用户的意图和手势来进行研究。这种研究方法能确定出不同用户错误率范围下，照片排序程序在不同的实验设置下得到的每个对话回合的平均收益值。图 8 显示了这些模拟的结果，可以看出，基于图 5 所

描述的经典算法的 手工策略 在低错误率时性能很好，但是在高噪声时会导致性能的显著下降。图8中，其他的曲线分别对应于由强化学习训练得到的各种策略。标记为 固定策略(30% 噪声) 的曲线给出的是采用错误率为30%的数据来训练的观察特征矩阵模型参数和最优策略的系统性能。可以看出，相对于 手工策略 ，该策略在高噪声时，鲁棒性有所提高，但是在低噪声性能受损。通过检查这个策略，我们发现该策略在低错误率时过度谨慎，总是选择 什么都不做操作 ，由此浪费了很多操作机会。标记为 固定策略+自适应模型 的曲线描述的系统采用了与前面相同的策略，但观察特征的概率矩阵会根据实际的错误率进行调整。可以看出，低错误率时的性能现已恢复，高错误率时的性能得到了进一步的提高。这显示了精确的模型参数的重要性。最后，标记为 自适应策略+自适应模型 的曲线给出了当策略也根据错误率进行自适应调整时的性能。在这种情况下，进一步改善了性能。



[图8] 4种结构每轮对于手势错误率的平均收益值。这4种结构为：手工策略、在30% 噪声下训练的固定策略、具有自适应模型参数的不变策略和具有自适应模型参数的自适应策略

总之，这些性能结果显示出了贝叶斯置信跟踪和策略优化对不确切和模糊的手势具有鲁棒性。图8中所示系统的性能提高有三个主要原因。首先是采用状态转移概率模型来描述环境的变化使用户的行为特征可以被用来消除手势中的歧义。第二个是采用显示的观测特征概率模型可以对噪声特性进行建模，这样可以优化隐式决定的阈值。第三,强化学习使得最优策略可以最大化收益值的期望，也即优化对话目标。

当然这只是一个用以说明基本的思想的小例子。图8中所示的性能和测试结果应谨慎对待。例如，在存在自适应的情况下，用户模拟器使用的参数与系统的参数完全相同，两者是完全匹配的。因此，图8中上部的曲线代表的是一个上限，该上限在实践中是难以实现的。此外，设计 手工策略 时没有使用收益值函数的知识，因此，使用平均收益作为性能测度是有利于基于数据训练的系统的。但无论如何，将贝叶斯置信状态跟踪与通过强化学习进行策略优化结合在一起的潜在技术价值还是非常清楚的。

正如引言所述，本节概述的系统是一个POMDP 的例子。POMDP 满足了认知型用户界面所需的所有条件：它们支持基于贝叶斯置信状态跟踪的推理和推论；它们采用优化的策略在不确定性条件下进行规划，这些策略是基于置信状态并通过强化学习训练得来的；它们是参数化的模型，从而可以迅速的进行在线自适应；因为策略是通过数据训练的，所以它们可以在更长的时间范围里从经验中进行学习和更新。

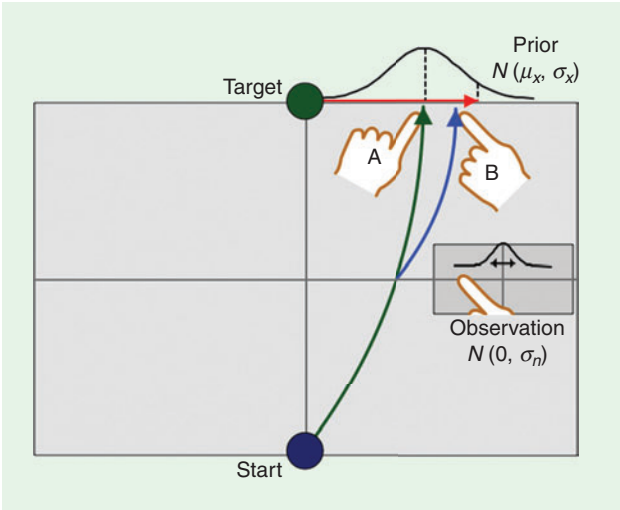
POMDP 绝非最近才出现。它们最初出现在运筹学的研究中[2][5]。机器学习领域的研究者已对其进行了广泛的探讨。但是它们的广泛使用却遇到了可解性方面的严重阻碍。较早的时候，当有人指出置信跟踪和策略优化的复杂度对状态空间的规模呈指数级变化时，可解性问题就已经被意识到了。事实上，它们的复杂度对操作和观察特征空间的规模也是指数级的增长关系。因此它们在现实世界中的应用并不简单，这点在后面还会被提及。但在此之前，让我们先来看看人类是如何在不确定条件下进行策略规划的。

人类决策和规划

本文的中心原则是：未来人机界面需要表现出认知的能力，这样才能满足下一代计算机系统要实现的目标，而贝叶斯推理和强化学习则必须用来支撑这种界面的实现。大多数互动是人和机器之间的一种协作行为。如果你知道在人的这一方面其实也遵循和机器类似的机制，你应该会感到欣慰吧。其实，人类采用强化学习的原则是不言而喻的[9]，但人类是否具有贝叶斯推理的能力就不那么显而易见了。所以我们提出这样一个有趣的问题 人类的决策是否具有贝叶斯统计的特性？

从进化角度说，脑功能发展的主要动力之一是运动。事实上，可以说人类有大脑的唯一原因是使他们可以移动[11]。因此要了解人类推理的核心机制，就非常有必要了解人类是如何规划自己的运动的。为回答这一问题，研究者们已经做了许多实验，但是第一次以解决运动规划是否具有贝叶斯特性的问题的，是我剑桥大学的同事 Daniel Wolpert。

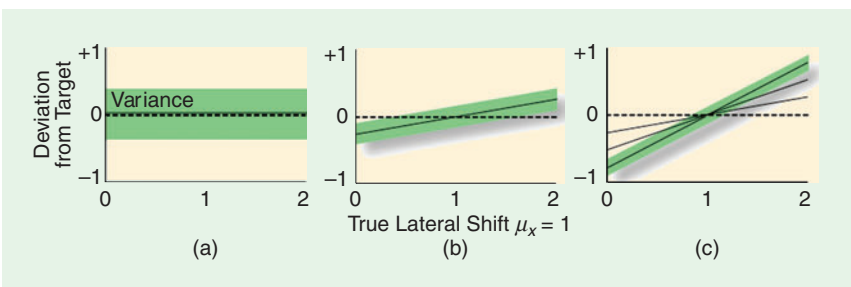
沃伯特的实验从概念上来说很简单，主要原理如图9所示[12]。要求一个测试者将他的手指穿过一张桌子，从蓝色的起始点到绿色的目标点。不过，在移动过程中，测试者的视线被遮挡，而参考框架



[图9] 一个简单的运动规划的任务。测试者必须将他的手指从蓝色的“开始”移动到绿色的“目标”。不过，在运动过程中，这个测试者的视线被遮挡，参考框架是沿高斯分布 $N(\mu_x, \sigma_x)$ 上的采样点 x 移动的。然后重复这一试验，测试者学会将她的手指移动到高斯分布的均值（如手A所示）。之后，测试者在从起始点到目标点中的途中允许向手指模糊的看一眼，则测试者将修改他们的轨道，如手B所示。接下来的问题是测试者“使用何种模型以做出正确的选择”？

则会偏移一段距离，这个距离 x 是从高斯分布 $N(\mu_x, \sigma_x)$ 上采样得到的。在每个操作结束，测试者可以去看她错过的距离，重复几次训练后，测试者学会将她的手指移动到高斯分布的均值 μ_x（图9中的路径 A）。然后这个实验被重复一遍，但在这次实验中，测试者在从起始点到目标点中的途中允许她对手指相对于参考点的位置有模糊的一瞥。此处的模糊相当于对观察加入了高斯噪声 $N(\mu_x, \sigma_x)$（图9中的路径 B）。这样经过多次试验，测试者至少在原则上已经有了一个先验的目标点位置信息和一个有噪声的观测结果。接下来的问题是测试者是如何使用此信息来调整她的运动规划。

在给定运动中途的带噪声的观察结果的条件下，测试者可能使用三种模型去确定目标的位置。



[图10] 三种模型偏离目标的平均预测偏差。(a) 全额补偿；(b) 直接映射；(c) 贝叶斯. 带厚度的绿色带表明差异变化。(c) 中曲线的坡度随观察噪声 的增加而增加。

首先，可以忽略先验信息，只使用带噪声的观察值来预测目标。在这种情况下，如图10（1），平均误差为零，但会有一个如绿色带所示的大的方差。

第二，测试者可以学习噪声观察和随之而来的目标错位之间的直接映射。通过最大限度地减少大量实验中的误判，测试者可以学习到某种最优映射而且无需明确使用先验分布或观测噪声。由于观测噪声的实现是通过将目标图像进行模糊化而产生的，测试者可以从视觉上估计 σ_n。然而，在实验中，他们只是在模糊度为零的情况下（其中 $\sigma_n = 0$）看到过目标的偏移错位。因此，如果他们使用直接映射算法，那么对于所有的试验，不管偏移错位的实际值是多少，他们都不得不使用相同的 $\sigma_n = 0$ 时的映射。这导致了测试者的行为反应如图10(b)所示。最后，假设人类将先验分布和观察分布内在化，则贝叶斯规则可以用来预测目标。这将导致最大后验估计：

$$x = \frac{\sigma_n^2}{\sigma_n^2 + \sigma_x^2}\mu_x + \frac{\sigma_x^2}{\sigma_n^2 + \sigma_x^2}x \qquad (5)$$

从图10(c)可以看出，平均标准差会随偏移的大小而改变，而斜率则取决于观察噪声的变化。注意，这种模型显示出的方差是最小的，而事实上，这正是这个估计问题的最小方差解决方案。

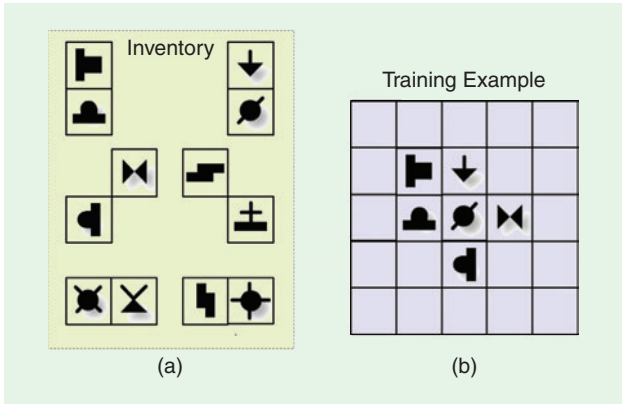
这个实验的结果表明毫不含糊地说明贝叶斯模型是唯一适合实验数据的模型。对数据的进一步分析显示测试者确实在学习先验分布。更进一步，当采用一个双峰高斯分布用于先验概率的时候，同样获得了一致的结果[13]，[14]，这说明人类可以预测和计算比简单的高斯分布更加复杂的分布。

最近的研究表明人类也将贝叶斯推理用于其他的处理活动。例如另一剑桥的同事，Máté Lengyel，已经证明人类将贝叶斯学习用于视觉分块[15]。在他的实验中，测试者要去观察如图11(b)所示的图形模式。这些模式是由图11(a)中所示的基本的积木块拼接组合而成的。这些基本积木块测试者是不知道的。经过训练后，一系列新的积木组合被拿给测试者，这些组合物中的一些来自测试积木清单，而另一些则不是。

当测试者被询问，在每一种情况下，基本的积木块是否与他看到的积木组合物相似的时候，他们通常会以75%左右的概率认为从测试清单中的积木是相似的，而这远高于随机的概率。对于人类如何做到这一点以及相关的复杂学习算法，研究者们已经

提出了若干模型。与它们相比，Lengyel 提出，人类会自然使用贝叶斯分块学习过程，包括利用奥卡姆剃刀原理来确定最优模型复杂度及积木块大小。通过改变组合物出现的频率和组合的复杂度（如使用3个），这些假设的机制可以通过数据进行仿真测试进而和人类性能进行比较。在所有情况下，贝叶斯方法和数据最匹配。

总体而言，实验数据表明，人类可以隐含的计算贝叶斯统计量，并使用贝叶斯推理来解决不确定条件下的规划问题。这里介绍的经验证据后来被许多进一步的实验证实[16]–[18]。还有一些生理架构上的论据也支持这种猜想，即人类的神经系统非常适合贝叶斯推理[19]。因此，似乎很清楚的是，人类已经进化出一种能力：既能通过观察量来学习先验的统计分布，又能利用贝叶斯公式去从这些分布中推理出后验分布。由此看来，人类确实在解决问题和在不确定条件下的规划时使用了贝叶斯推理。



[图11] 视觉组块任务：测试者表现为表格的训练模式如(b)所示，积木目录如(a)所示，目录保持隐藏。让测试者观看目录中的个别积木，并问他对积木是否熟悉。

到现实世界系统的扩展

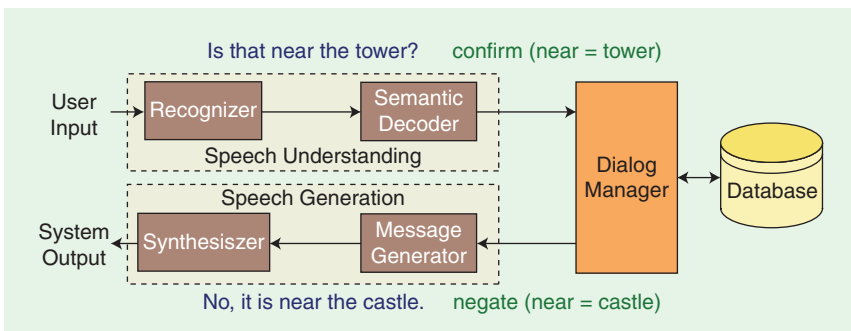
在 例子：一个简单的触摸手势驱动界面 这一节中，给出了POMDP 的基本思想，并通过一个简单的例子说明了它们有潜力实现鲁棒和智能的用户界面。POMDP 框架的主要特征是：存在一个可以持续运行的由置信状态组成的系统，置信状态的更新通过贝叶斯推理实现，系统策略的性能可以用收益值量化衡量，策略的优化通过强化学习来实现。如前一节所述，有充分的证据显示人类也在利用类似的机制。总而言之，POMDP 似乎满足前面所述的认知型界面的所有要求。那么，为什么POMDP框架并未用于目前使用的用户界面呢？

对这个问题的答案在 例子：一个简单的手势驱动界面 这一节结束时已经提到。一个现实世界中的人机界面的状态空间的规模是巨大的。因此，通过（1）的置信跟踪算法实现的话，对于实时系统来说成本太高了。此外，POMDP策略的精确实施和优化往往都是难以处理的（玩具级的问题除外）。不管怎样，这些问题都不必然是取得进展的障碍。

POMDP框架的基本要素是：使用多个可能值来描述不确定性，以及采用可优化可量化的决策过程。有几种POMDP的近似算法可以保持这些基本要素，并在实践中取得了比传统方法好得多的性能。本节的剩余部分将通过一个统计对话系统（Spoken Dialogue System: SDS）的设计来对这些加以说明。

SDS被广泛用于银行、金融和交通领域的语音信息服务。最近，它们也越来越多的被用于自动呼叫中心。用于旅游信息查询领域的一个典型的口语对话系统结构如图12所示。用户的声音先通过一个语音识别系统转换成文字，然后通过语义识别器将这些文字转换成 对话行为 。 对话行为 是对用户意图的一个抽象化的描述，例如inform(food=chinese)，confirm(near=tower)。一般情况下，对话行为的 类型 与对话系统的实际应用领域是无关的，而 属性-取值 则与具体应用有关。用户的对话行为传递给对话管理器。该对话管理器可以解读输入的用户对话行为，更新其内部状态，并且以对话行为的形式产生输出反馈。这个系统输出的对话行为会被转化为自然语言，并由语音合成器合成语音。

SDS包括了设计现实世界中的认知型用户界面过程中会遇到的所有问题。系统内部状态s通常可分解为三个要素：$s = \{g, u, h\}$，其中 g 代表用户的意图，u 代表用户输入的对话行为，h 代表对话历史[20]。所有这些都是极为复杂的。此外，由于语



[图12] 用于旅游信息域的一个口语对话系统的体系结构。该语音识别系统产生一个词串，语义识别器将其转换为一个称为对话行为的用户意图一个抽象表示。用户的对话行为传递给一个对话管理器，管理器解释对话行为，更新它的内部状态，并以输出对话行为的形式产生一个适当的反应。然后将对话行为转换为自然语言进而由语音合成器转换成语音。

音识别错误率通常很高，通过识别得来的用户输入 u 存在很大的不确定性，这些不确定性也会传播到 g 和 h 中去。再有，由系统的各种操作组成的空间必须涵盖所有可能的系统响应操作，对话管理器的策略必须能将复杂和不确定的对话状态映射到巨大的操作空间中。所有这些因素结合起来，使 POMDP 框架内 SDS 的实现成了一项重大的挑战。

尽管如此，通过利用一些简单的想法，POMDP 框架还是可以扩展到现实世界中。首先，通过简化对话状态概率分布的表示形式可将置信跟踪变得易于处理。例如，在旅游信息查询的应用中，用户意图包括四个离散值：类型，位置，价格和食物。

精确的置信跟踪需要知道这些变量的完整的联合概率分布，P(类型, 位置, 价格, 食物)；但即使是很有限个数的类型，地点，价格点和食物种类也会使整个联合概率分布的规模大到不可想象。处理这一问题的最简单方法是使用 $M-best$ 近似，即对所有意图状态值的概率进行排名，只保留了 M 个最可能的状态，其余删除。例如，旅游信息查询的应用中可能出现如下例子：

P(旅馆, 东方, 便宜, 无) = 0.65

P(旅馆, 西方, 便宜, 无) = 0.21

P(饭店, 东方, 便宜,意大利的) =0.08

P(酒吧, 东方, 便宜, 无)=0.04

P(旅馆, 东方, 贵, 无)= 0.01

除了以上状态组合之外，其余的各种组合的概率都太低，不需要保留。

对于置信状态的第二种近似方法是：通过设定一些独立假设来分解整个联合分布。例如，从旅游应用的特点出发，我们有理由认为食物和场所（例如餐馆）的价格仅仅依赖于其类型，而类型和位置则是相互独立的，即：

$$P(\text{类型, 位置, 价格, 食物})$$

$$\approx P(\text{价格}|\text{类型})P(\text{食物}|\text{类型})P(\text{类型})P(\text{食物}) \quad (6)$$

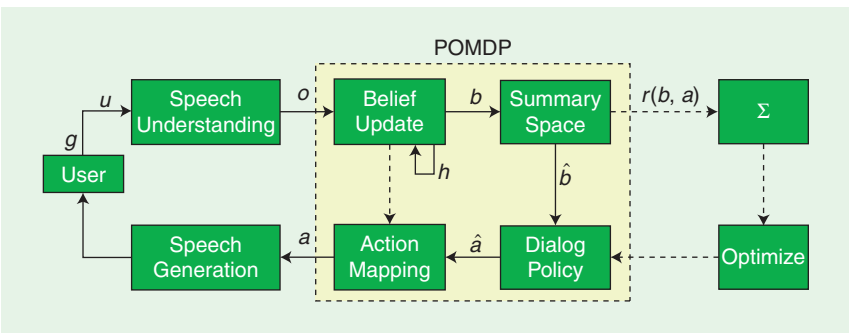
这就给出了意图状态的一个贝叶斯网络表示。

上述两种近似方法都可以使置信跟踪易于处理，但它们没有改变状态空间过大以至于无法进行有效决策优化的问题。处理这类问题的通常办法是通过映射函数 将所谓 主置信空间 b 映射到更紧凑的 摘要置信空间 $\hat{b}$。以同样的方式，可以在摘要空间中定义一个紧凑的操作集，而这个紧凑操作集可以通过逆映射 再映射回主操作空间[21]。

下面的两节将概述两个在剑桥大学实现的统计对话系统，它们分别代表了以上两种不同的 POMDP 框架下的近似方法。第一个是采用 $M-best$ 近似算法的HIS系统。第二个是采用贝叶斯网络方法的 BUDS 系统。二者都使用了从主空间到摘要空间的映射，但是方式有所不同。[22]给出了统计对话系统的更多详细总结。

隐信息状态系统

HIS系统的框图如图13所示[23][24]。它对置信跟踪过程采用 $M-best$ 近似，同时在策略优化中采用了摘要空间映射技术。图12给出了一个典型对话回合的基本流程。用户输入通过语音理解模块处理，该模块输出一个列表，包括N个最好候选值以及相关的置信度[<u1, c1> , , <uN, cN >]。我们将整个列表视为POMDP对话管理器得到的观测值 o，据此更新置信状态 b，然后将更新后的置信状态映射成为摘要置信状态 $\hat{b}$。而对话策略则通过 $\hat{a}=\psi(\hat{b})$ 将摘要操作与每一个可能的摘要置信状态 $\hat{b}$ 联系在一起。然后摘要操作可以被映射回主操作空间，形成系统的操作响应 a。



[图13] HIS系统. HIS的对话管理器保持着对所有可能的对话状态置信分布。为了使策略具代表性和易于优化，将置信分布映射到一个简单的摘要空间。通过使用一个启发式操作映射将摘要操作扩展到整个系统操作向用户产生反应

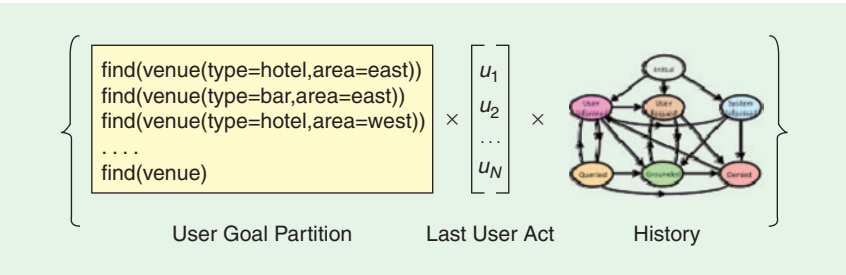
HIS 系统由前面所述的三个要素构成，即其中 g 代表用户意图，u 代表用户输入的对话行为，h 代表对话历史。如果将其按（1）式分解，给定合理的独立假设，易见：

$$b'(g',u',h') = k \cdot \underbrace{P(o'|u')}_{\substack{\text{oberservation}\\ \text{model}}} \underbrace{P(u'|g',a)}_{\substack{\text{user action}\\ \text{model}}} \sum_{g,h} \underbrace{P(g'|g,a)}_{\substack{\text{observation}\\ \text{model}}} \underbrace{P(h'|g',u',h,a)}_{\substack{\text{dialogue distory}\\ \text{model}}} b(g,h) \quad (7)$$

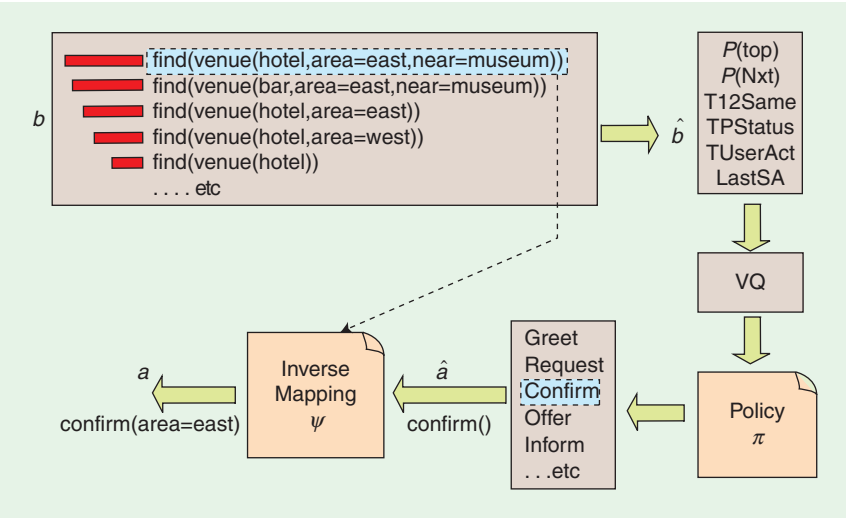
其中，　　符号表示下一个时间点[25]。如底部括号所示，对话系统的置信状态更新方程涉及四个不同的概率模型。用户意图模型（user goal model）和对话历史模型（dialogue history model）表示了马尔可夫决策过程的运行方式。在HIS系统中，假设用户意图不会改变，对话历史模型由一个确定的具有有限状态的完毕模型代替。更有趣的是观测模型（observation model）和用户操作（user action model）模型。观测模型包含了语音理解系统的误差信息，它是前面iPhone例子中观察特征概率矩阵的一个推广。而用户操作模型实现了对观测模型的概率进行缩放。由于观测值是包含N个最好候选值的列表，用户操作模型实际的效果是根据上下文对这个列表进行了重新排序。因此，用户操作模型提供了上下文敏感的过滤器，它能非常有效的减少失误，特别是在高噪声条件下[26]。

为了进一步简化置信跟踪，HIS系统将用户意图状态分为一系列等价类，称为　分区　。在对话开始时，所有意图状态都在一个分区。通过观测到的用户对话行为，新的证据不断被累积，分区也据此被不断细分以描述不同的用户意图。这种细分过程会遵循一套从数据库中导出的本体规则，这些规则具有树状过程树状结构，以确保所有分区的并集始终等于完整的状态空间。由于一个分区中所有的意图状态在当前证据下无法进一步被区分，因此置信状态的更新就可以在分区层次进行，而无须对每个单独的意图状态进行，这就大大减少了计算量。HIS的状态空间生成的示意图如图14所示。每一个HIS分区包括：一个用户意图分区，前一次观测到的用户对话行为列表中的一项，以及完毕信息。完毕信息构成了对话历史，以使得这些分区中的每个树节点都有一个根据确定的有限状态机规则变化的完毕状态。因此总体来说，HIS的状态空间由所有可能的分区加上所有可能的用户对话行为和所有可能的完毕状态组合而成。这个集合中每个状态分区的概率会被计算、排序或者剪枝。系统通常会维持300-3,000个活跃分区，所有这些活跃状态组成了系统的置信状态b。

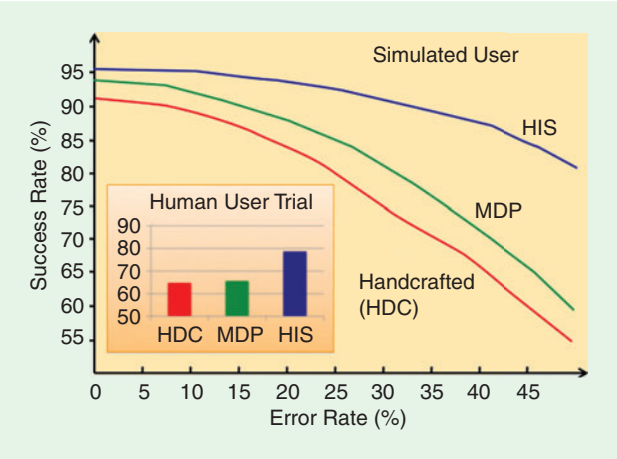
主空间到摘要空间映射和HIS系统的策略表示见图15。摘要置信状态$\hat{b}$由固定维度的特征矢量组成，例如主空间中最优状态的概率和次优状态的概率，加上一个布尔变量，表示这两个状态是否可以用来代表同一个实体项。摘要矢量被映射到经过矢量量化的摘要空间中的一个固定网格格点上[27], [28]。每个格点都有一个对应的摘要系统操作，这个操作通过逆映射函数可以映射回主操作空间，在大多数情况下，我们都假定该系统操作的主题是由主置信状态空间中的最优状态决定的。这个过程如图15所示，　确认　这个操作类型是从摘要空间中选择的，而需要确认的主题内容（即属性值）则来自于置信状态b中具有最高概率的那个状态。



[图14] HIS置信空间. 每一个用户意图分区，前一个用户对话行为，和历史状态的集合构成HIS分区空间。历史状态记录了完毕信息，其详情请见[24]。



[图15] HIS摘要空间映射。摘要空间$\hat{b}$由一个固定维数的特征矢量组成，通过矢量量化器映射到摘要空间中的一个固定网格点。每个格点都有一个对应的摘要系统操作，该操作被选择并且通过逆映射函数将其映射回主空间。



[图16] 作为输入错误率函数的HIS的性能。主要的图形显示成功率的比例随使用一个模拟用户的语义错误率函数的变化。该插图显示了类似结果，为人类用户在噪声条件下进行了该试验，其平均误差率为25%。

连续摘要空间的矢量量化将 HIS POMDP 转换成一个简单的离散马尔可夫决策过程（MDP），对此，存在许多优化算法。HIS系统就是采用了蒙特卡罗控制算法，在与用户模拟器的交互过程中，通过在线的强化学习来估计最优的系统操作集合[29]。

图16显示了HIS系统的性能，与只保持一个最优对话状态的基于传统马尔科夫决策过程（MDP）的对话管理器相比，它保持多个候选状态，但是只有一个手工定制的策略。上图中的曲线给出了平均成功率随着用户模拟器输入语义错误率的函数变化。 成功 的定义是：系统给出了满足用户需求的场所，并提供了用户要求的任何信息，如地址或电话号码。可以看出，HIS系统对于高错误率更具鲁棒性。柱状图则给出了在噪声环境下与36个真人测试者进行对话的性能。HIS系统明显更具鲁棒性。

对话状态系统的贝叶斯更新

前一节所述的 HIS 系统显示了利用置信空间的 $M-best$ 近似算法，再加上摘要状态映射可以实现复杂的现实世界对话系统。虽然相对于传统系统，HIS 系统能够提供更高的性能，但它却有两个主要问题：首先，$M-best$ 近似方法使其很难使用状态转移矩阵，因此，HIS 系统假定在对话过程中用户意图不会改变。第二，HIS 系统中的概率模型是确定性规则与统计模型的混合，很难完全通过数据自动训练。

正如前面介绍的，我们可以采用另一种近似：用贝叶斯网络表示对话状态。这种近似保留了正确反映系统动态变化的能力，并可以充分利用参数化的模型；但忽略了现实世界中固有的很多条件相关性。

这种方法的一个例子是在剑桥建立的称作BUDS的系统。该系统使用了与HIS系统同样的对话状态分解 $s(g,u,h)$，但它却将每个组分进一步分解成 概念 。例如，在旅游信息领域，用户可能会对位置，价格，食物，星级和音乐等概念感兴趣。这些概念大部分将依赖于所涉及的场地类型（餐厅，酒吧，宾馆等），但除此之外，它们可以被视为是独立的。图17给出了一个动态贝叶斯网络结构，它给出了场地类型和一个相关的概念：食物。请注意，在实际系统中，根据不同的应用，有10到20个概念。每个概念 C 有三个节点：一个意图节点 g_c，其值随用户可能的选择变化；一个用户操作节点 u_c，表示的是前一个用户对话行为的类型，如果上一个用户操作并没有提及这个概念，则该行为为空；一个历史节点 h_c，取值为一个简单的完毕模型值，如（初始化，涉及到，完毕）。所有的 u_c 节点都依赖于一个描述完整的用户对话行为的节点，于是它们也必然取决于观测值，这里的观测值与 HIS 系统一样，是一组用户对话行为的 $M-best$ 候选列表。系统的动态变化可以通过使当前节点与前一时间点中的等效节点相关联来实现。在 BUDS 系统中，意图节点和历史节点都与于它们以前的值相关。

BUDS系统将所有相关的对话状态信息都表示在一个贝叶斯网络结构中。于是置信跟踪就可以利用任何现有的贝叶斯网络近似推理算法来实现。在BUDS系统中，我们使用的是循环置信传播(LBP)。然 而，为使其能实时运行，各种优化还是必要的。例如，意图节点的值域可能是一个很大的集合，而在单次对话中，一般只会涉及极少的一部分。因此，与在HIS系统中的做法相似，我们对这些值进行分区，这样，LBP 会运行得更快。这通常能将有效基数下降到二或三，这对降低计算时间有很好的效果。另一种非常有效的优化是假设用户意图不断变化的概率是恒定的。通过降维，对于一个用户意图转移矩阵规模为 n 的问题，时间复杂度从 $O(n^2)$ 可以降低到 $O(n)$

将置信空间分解为在大量的因素之后，系统策略就要采取同的表示方式，因为将每个主状态映射到摘要空间已不再可能。在 BUDS 系统中，我们采用一种 SoftMax 形式的随机策略，其参数为 θ，表示如下：

$$\pi(a\,|\,b,\theta)=\frac{e^{\theta.\phi_a(b)}}{\sum_{a'}e^{\theta.\phi_{a'}(b)}} \qquad (8)$$

其中，$\phi_a(b)$ 是操作 a 的基函数。这些基函数可进一步分解成分量，使贝叶斯网络中的每个概念意图节点都会对整体策略发生影响

$$\phi_a(b)^T = [\phi_{a,1}(b)^T, ..., \phi_{a,G}(b)^T, \phi_{a,*}(b)^T] \quad (9)$$

其中，下标$1...G$在意图节点范围内变化，最后的一项$\phi_{a,*}(b)^T$包括全局信息，例如有多少数据项满足用户最可能的目标意图。这个参数化的策略可以通过最大化期望收益值来优化。我们发现，BUDS 系统中，natural actor critic算法是非常有效的[32]。

与 HIS 系统一样，BUDS 系统已经在用户模拟器和真人测试环境中都进行了实验。性能结果相同或略好于 HIS 系统结果，在此不再赘述。但是 BUDS 系统的主要优点是，可以采用Dirichlet 先验模型来把系统模型参数本身也引入到贝叶斯网络中去。如果用期望置信传播代替循环置信传播[33]，该系统就可以从数据中在线学习模型参数并进行自适应。因此与 BUDS 相似的架构可以满足前面所述的认知型用户界面的所有要求。

结论与展望

本文主要讲述以未来的计算机系统所需要的用户界面，它将支持比当前方法更具鲁棒和智能的交互。本文认为，未来的界面必须提供认知功能，即具有支持推论和推理的能力、在不确定性条件下进行规划的能力、短期适应的能力和长期从经验中学习的能力。满足认知型用户界面要求的工程框架应基于 POMDP，POMDP结合了贝叶斯置信跟踪和基于收益值的强化学习。实验证明，这个框架可以对不精确的人类交互信息进行鲁棒的理解，同时有能力通过最大化目标函数来搜寻最优的交互策略。更进一步，人类本身其实也是利用了类似的机制。

如果这个观点被接受，则其影响是：我们必须要以一种不同的方式来设计以人为中心的IT系统。关键是要确定哪些是不确定性的主要来源，及他们如何能有效地在系统内表示。用户输入必须被当作证据，据此，通过贝叶斯推理，不确定性可以得到解决。虽然 POMDP 往往被认为在现实世界应用中存在可解性的问题，但是，在实践中合理使用近似算法就仍然可以实现实际系统，并且同时保留 POMDP 框架的基本优势。

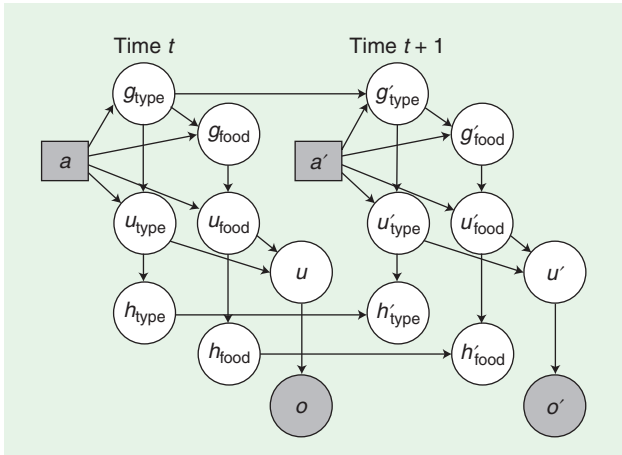
文章中详细描述了 HIS 口语对话系统的实现。HIS 系统同时代表了从现有传统系统到上述新系统的一个演化路径。实际从效果上看，HIS 系统等价于若干个对话管理器在并行运行，其中每个对话管理器对用户意图有各自不同的假设。HIS 系统包含很多符号化的组件，这与传统对话系统很相似。而事实上，将传统的对话系统模块集成到一个概率框架中去正是它的优点。

然而，从长远来看，认知型用户界面只有设计成彻底的概率模型，才能保证系统能够随时间进化且能适应经验。对于此类系统，毋庸置疑可以采用贝叶斯网络(如 BUDS 系统)。然而，基于大规模贝叶斯网络的 POMDP 系统的实现面临很多的挑战。最直接的问题就是当网络复杂性增加时如何保持实时性操作，这是由于近似推理需要大量的计算。正如 BUDS 系统所展示的，应用传统的置信传播算法，以及人类交互通常把重点放在少量具体实体上这个基本事实，显著的速度提高是可以达到的。但是，最终我们仍需要能处理大规模的动态变化系统，而这也许需要底层硬件的支持，也许需要某种特殊的分布式处理器用来优化置信传播算法所需的消息传递操还作。也有其他的挑战，例如集成多通道输入和输出，以及处理人和人自然交流之间的微妙对话现象。还存在一些社会问题，与现存的其他系统不同，我们可能无法精确保证一个认知型用户界面在一些特定情况下将如何反应。

除了少数明显的例外，IT 系统的传统设计思路是将所有的信息列成表，再采用确定性算法来操作。输入也被看成确定性的，这在语音识别技术的应用上带来了明显的困难：语音识别被作为键盘输入的替代品，要使它有用就必须要降低错误率。本文认为，这种看法是错误的。本文认为真正的认知型人机界面会需要一种全新的，以对不确定性进行明确建模为核心的方法。POMDP 为合理设计此类系统提供了一个良好的基础框架，它是未来认知型用户界面的关键。

致谢

笔者感谢剑桥对话系统小组的现任和前任成员：Milica Gašic´、Filip Jurc˘íc˘ek、Simon Keizer、Fabrice Lefevre、François



[图17] BUDS使用动态贝叶斯网络，其中，将对话状态分解为代表的诸如“食物、价格和位置”等手势插槽。每个槽有目标g，一个相关的用户对话行为u和历史信息h。除了依赖于槽场地类型的插槽外，该槽大多是独立的。

Mairesse、Jost Schatzmann、Matt Stuttle、Blaise Thomson、Karl Weilhammer、Jason Williams 和 Kai Yu。本文中提到的一些研究是由英国 EPSRC（资助协议EP/F013930/1）和欧盟FP7计划（资助协议216594）资助。(CLASSIC项目：www.classic-project.org). 他们也给出了匿名的评论和建议，对改善本文的最后版本颇有帮助。

作者简介

Steve Young (sjy@eng.cam.ac.uk) 现为剑桥大学副校长，剑桥大学工程系信息工程分部教授。他主要的研究兴趣在于口语系统，包括语音识别，语音合成，语义理解，统计对话管理。他是 HTK Toolkit 的原创者，和 Phil Woodland 一起，开发了 HTK 词汇语音识别系统。参考文献1993年到2004年间，他担任 Computer Speech and Language 的编辑，目前是 IEEE 语音和语言处理技术委员会的主席。他是英国皇家科学院院士，英国工程技术学会、IEEE 和 RSA的会士（fellow）。2004年，他获得了IEEE信号处理学会技术成就奖。2008年，他当选为国际语音通信协会（ISCA）的会士，2010年他获得了ISCA的科学成就荣誉奖章。[SP]

参考文献

[1] Apple Inc. (2009). iPhone human interface guidelines [Online]. Available: http://developer.apple.com/iphone/library/documentation

[2] E. Sondik, "The optimal control of partially observable Markov decision processes," Ph.D. dissertation, Stanford Univ., Palo Alto, CA, 1971.

[3] L. Kaelbling, M. Littman, and A. Cassandra, "Planning and acting in partially observable stochastic domains," Artif. Intell., vol. 101, pp. 99–134, 1998.

[4] C. Bishop, Pattern Recognition and Machine Learning. New York: Springer, 2006.

[5] R. Smallwood and E. Sondik, "The optimal control of partially observable Markov processes over a finite horizon," Oper. Res., vol. 21, no. 5, pp. 1071–1088, 1973.

[6] R. Sutton and A. Barto, Reinforcement Learning: An Introduction (ser. Adaptive Computation and Machine Learning). Cambridge, MA: MIT Press, 1998.

[7] M. Littman, A. Cassandra, and L. Kaelbling, "Learning policies for partially observable environments: Scaling up," in Proc. 12th Int. Conf. Machine Learning, A. Prieditis and S. Russell, Eds. San Francisco, CA: Morgan Kaufmann, 1995, pp. 362–370.

[8] Y. Virin, G. Shani, S. E. Shimony, and R. Brafman, "Scaling up: Solving POMDPs through value based clustering," in Proc. 22nd Nat. Conf. Artificial Intelligence, AAAI 2007, Vancouver, 2007.

[9] W.-T. Fu and J. Anderson, "From recurrent choice to skill learning: A reinforcement-learning model," J. Exp. Psychol. Gen., vol. 135, no. 2, pp. 184–206, 2006.

[10] G. Cziko, Universal Selection Theory and the Second Darwinian Revolution. Cambridge, MA: MIT Press, 1995.

[11] D. Wolpert, Z. Ghahramani, and J. Flanagan, "Perspectives and problems in motor learning," Trends Cogn. Sci., vol. 5, no. 11, pp. 487–494, 2001.

[12] K. Kording and D. Wolpert, "Bayesian integration in sensorimotor learning," Nature, vol. 427, pp. 224–227, 2004.

[13] R. Jacobs, "Optimal integration of texture and motion cues to depth," Vision Res., vol. 39, pp. 3621–3629, 1999.

[14] M. Ernst and H. Bulthoff, "Merging the senses into a robust percept," Trends Cogn. Sci., vol. 8, pp. 162–169, 2004.

[15] G. Orban, J. Fiser, R. Aslin, and M. Lengyel, "Bayesian learning of visual chunks by human observers," Proc. Nat. Acad. Sci., vol. 105, no. 7, pp. 2745–2750, 2008.

[16] K. Kording and D. Wolpert, "Bayesian decision theory in sensorimotor control," Trends Cogn. Sci., vol. 10, no. 7, pp. 319–326, 2006.

[17] H. Tassinari, T. Hudson, and M. Landy, "Combining priors and noisy visual cues in a rapid pointing task," J. Neurosci., vol. 26, no. 40, pp. 10154–10163, 2006.

[18] D. Wolpert, "Probabilistic models in human sensorimotor control," Hum. Mov. Sci., vol. 26, pp. 511–524, 2007.

[19] M. Sahani and P. Dayan, "Doubly distributional population codes: Simultaneous representation of uncertainty and multiplicity," Neural Comput., vol. 15, pp. 2255–2279, 2003.

[20] J. Williams, P. Poupart, and S. Young, "Factored partially observable Markov decision processes for dialogue management," in Proc. 4th Workshop Knowledge and Reasoning in Practical Dialogue Systems, Edinburgh, 2005.

[21] J. Williams and S. Young, "Scaling POMDPs for spoken dialog management," IEEE Trans. Audio, Speech and Lang. Processing, vol. 15, no. 7, pp. 2116–2129, 2007.

[22] O. Lemon and O. Pietquin, "Machine learning for spoken dialogue systems," in Proc. Interspeech, Antwerp, Belgium, 2007, pp. 2685–2688.

[23] S. Young, J. Schatzmann, K. Weilhammer, and H. Ye, "The hidden information state approach to dialog management," in Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing, ICASSP 2007, Honolulu, HI, 2007.

[24] S. Young, M. Gasic, S. Keizer, F. Mairesse, J. Schatzmann, B. Thomson, and K. Yu, "The hidden information state model: A practical framework for POMDP-based spoken dialogue management," Comput. Speech Lang., vol. 24, no. 2, pp. 150–174, 2009.

[25] J. Williams and S. Young, "Partially observable Markov decision processes for spoken dialog systems," Comput. Speech Lang., vol. 21, no. 2, pp. 393–422, 2007.

[26] S. Keizer, M. Gasic, F. Mairesse, B. Thomson, K. Yu, and S. Young, "Modelling user behaviour in the HIS-POMDP dialogue manager," in Proc. IEEE Workshop Spoken Language Technology (SLT'08), Goa, India, 2008.

[27] W. Lovejoy, "Computationally feasible bounds for partially observed Markov decision processes," Oper. Res., vol. 39, pp. 162–175, 1991.

[28] R. Brafman, "A heuristic variable grid solution method for POMDPs," in Proc. 14th Nat. Conf. Artificial Intelligence, AAAI, Cambridge, MA, 1997.

[29] M. Gasic, S. Keizer, F. Mairesse, J. Schatzmann, B. Thomson, and S. Young, "Training and evaluation of the HIS POMDP dialogue system in noise," in Proc. 9th SIGdial Workshop Discourse and Dialogue 2008, Columbus, OH, 2008.

[30] B. Thomson, J. Schatzmann, and S. Young, "Bayesian update of dialogue state for robust dialogue systems," in Proc. Int. Conf. Acoustics Speech and Signal Processing, ICASSP, Las Vegas, 2008.

[31] B. Thomson and S. Young, "Bayesian update of dialogue state: A POMDP framework for spoken dialogue systems," Comput. Speech Lang., to be published.

[32] J. Peters and S. Schaal, "Natural actor-critic," Neurocomputing, vol. 71, no. 7–9, pp. 1180–1190, 2008.

[33] T. Minka, "Expectation propagation for approximate Bayesian inference," in Proc. 17th Conf. Uncertainty in Artificial Intelligence, Seattle, WA, 2001, pp. 362–369.

[34] J. Henderson and O. Lemon, "Mixture model POMDPs for efficient handling of uncertainty in dialogue management," in Proc. 46th Annu. Meeting Association for Computational Linguistics (ACL'08), Columbus, OH, 2008.

[35] K. Kim, C. Lee, S. Jung, and G. Lee, "A frame-based probabilistic framework for spoken dialog management using dialog examples," in Proc. 9th SIGdial Workshop Discourse and Dialogue, Columbus, OH, 2008.

[36] T. Paek and R. Pieraccini, "Automating spoken dialogue management design using machine learning: An industry perspective," Speech Commun., vol. 50, no. 8–9, pp. 716–729, 2008. [SP]

Xiaodong He, Li Deng and
Wu Chou

语音识别中模型优化技术的一个综述

判别学习在序列模式识别中的应用

判别学习已经成为包含语音识别和语言处理在内的统计信号处理及模式识别研究领域的一个主题[9,10,13,21,29,35,43,44,47,49]。特别是近年来在大规模的语音识别中因为引入判别学习而取得了很多成就[35,38,47,48]。理解语音过程的重点是语音序列的可变长度的特性的动态描述。对于序列模式识别的判别学习有两个重要问题，一是建立优化目标函数，二是实际的优化技术。文献[9,18,21,29,35,38,41,45,52,54,58]为解决这两个问题提出了很多的方法。然而，它们并没有充分理解这两个关键问题之间的关系。由于该问题在理论和实际中均起到重要作

用，因此迫切需要去归类和统一文献中的判别学习方法。这篇文章旨在满足上述要求，同时对判别学习框架下的序列模式分类和识别提出一些见解。本文旨在说明这些判别学习方法之间的关联和区别，并利用深层次的方案去统一表面上不相同的各种技术。尽管本文综述了一般类型的关于序列特征的模式识别问题，但大部分重点还是放在讨论这类问题和语音识别及隐马尔科夫模型[11,50,56]之间 的关系上。隐马尔科夫模型和其他多种形式的判别学习一样已经被应用到很多的信号处理领域，除了语音领域，还有生物信息学[6,16]，计算遗传学[55]，文本和图像分类和识别[33,62,65]，视频目标分类[63]，自然语言处理[8,10]，遥控机器人[64]。我们希望通过本文的综述和见解可以使判别学习在应用更广的信号处理学科方面取得原理上的进步或者成功的应用。

除了给出大量基本的判别学习的思路和方法，也希望我们的相应算法从更一般的机器学习方法的角度出发，定位在更广泛的建立统计分类器的问题上。生成的方法(generative methods)和判别的方法(discriminative methods)是设计和训练统计分类器及识别器的两类范例模式。生成识别器依靠一个观测特征和相应的类别的联合概率密度的学习模型。使用这个联合概率模型，根据贝叶斯规则[12,50,66]计算最大后验概率完成决策任务。相反，判别分类器及识别器是直接使用类别后验概率（或者相关的判别方程），概括描述为“直接解决分类或识别的问题而不解决中间步骤的广义性问题”[57]。识别器的设计理念基础是广泛流行的机器学习方法，包括支撑矢量机[57]，条件随机场[32,44]和最大熵马尔科夫模型[19,34]，这里估计联合分布的中间步骤被省略。另一方面，尽管联合分布估计具有复杂性，其唯一目的就是判别，生成方法在对识别器的各个成分和交互方面的知识融合、概念直接分析有很重要的作用。

上述对这两种一般化的学习方法相应的优点和局限性的分析致使在这里需要一种实用的模式识别框架。也就是试图估计一种简单的联合分布或生成模型，它有比计算真实分布的生成模型更低的复杂度。为了使得低复杂性的生成模型判别足够好，这需要参数学习方法在实际的判别任务中克服简单模型结构的局限性。这与使用最大似然估计拟合内部数据的传统方法相反。这种框架已经在语音识别研究领域有了很重要的应用和指导意义，这里隐马尔科夫模型作为一种低复杂度的联合分布模型，常用于描述语音的局部声学特征序列和相应的底层语言标记序列（句子，单词或者音）的联合分布模型。典型隐马尔科夫模型的判别参数学习方法为：1)最大互信息(MMI)[7,18,20,39,40,41,58,61]；2)最小分类误差(MCE)[1,9,23,28,29,35,36,38,49,52,54]；3)最小音误差(MPE)和相近的最小词误差(MWE)[13,45-48]。

除了对上述分类方法做总回顾外，本文还在判别学习的三个关键领域给予特别关注。首先给出了三种用于分类器参数优化的主要判别学习目标函数(MMI,MCE,MPE/MWE)的统一概述，从他们的起源探究他们的关系。通过统一的目标函数，本文分析了对模式识别任务具有不同优化性能的各种条件，包括超字符串单元，字符串单元和子字符串单元的优化性能水平。其次，本文描述了一种在设计分类器时参数估计的有效方法。在判别学习中，这种参数估计方法基于增长转型(Growth Transform, GT)的优化框架，（在有理方程的最优化(Optimizing Rational Functions)中有很详细的介绍)。我们分析表明，该方法统一了参数估计公式，同时也可升级适用于大型的模式识别任务。本文第三部分说明了对于采用隐马尔科夫模型的序列模式识别问题，基于MCE和MPE/MWE的学习方法在增长转型的参数估计框架下的算法性质。

II. MMI,MCE,MPE/MWE的判别学习准则

MMI,MCE,MPE/MWE是在语音和语言处理领域三个最重要的判别学习准则。

尽管本文主要讨论语音和语言处理方面的判别分类器设计，但它们同样可以应用到其它相似的序列识别领域，如手写体识别。本文的参考文献涉及词语、音、字符串等的识别，就是为了说明序列动态识别问题可以基于不同层次的识别单元。此外，序列模式识别的分类器可以是基于每一个独立的模式或识别单元。如果可以利用序列的相关性，分类器的构造就可以基于字符串的模式或识别单元的识别，如短语，字符串，句子。该灵活性为序列模式识别的分类器设计提供了很大的研究空间，已经发展出很多的方法[22,29,47]。

首先Λ记为分类器参数的集合，在设计分类器时需要对其进行估计。在语音和语言处理中，对一个观察序列X，相应的标记词序列为S，其基于分类器的联合分布即为：

$$p(X,S|\Lambda)=p(X|S,\Lambda)P(S) \tag{1}$$

上式中，假设“语言模型”$P(S)$中的参数不需优化。给定一个训练数据集合，记R作为训练样本总数。本文主要讨论有监督的学习，这里每一次训练标记由一组观察数据序列$X_r=x_{r,1},\ldots,x_{r,T_r}$组成，其正确的模式序列标记为：$S_r=W_{r1},\ldots,W_{rN_r}$，其中$W_{r,i}$是序列$S_r$的第$i$个字。使用小写的变量$S_r$去记录所有可能的模式序列，这些序列可以用来标记第$r$个标记，包括正确的被标号序列$S_r$和其它序列。

A. 最大互信息(MMI)

在基于MMI的分类器设计方面，全局分类器参数估计是以数据X和相应的标号或者符号S之间的互信息$I(X,S)$最大化为目标的。从信息论角度看S和X的关系，信息量提供了信息获取量的一个度量，或者不确定性降低的数量。MMI准则在信息论中能够较好的估计。它具有很好的理论特性，同时又不同于用在基于生成模型的学习中的最大似然准则。互信息量$I(X,S)$的定义为：

$$I(X,S)=\sum_{X,S}p(X,S)\log\frac{p(X,S)}{p(X)p(S)}$$

$$=\sum_{X,S}p(X,S)\log\frac{p(S|X)}{p(S)}=H(S)-H(S|X) \tag{2}$$

其中$H(S)=-\sum_S p(S)\log p(S)$是$S$的熵，$H(S|X)$是给定数据$X$的条件熵：

$H(S|X)=-\sum_{X,S}p(X,S)\log p(S|X)$.。这里$p(S|X)$是基于模型$\Lambda$的，可以得到公式：

$$H(S|X)=-\sum_{X,S}p(X,S)\log p(S|X,\Lambda) \tag{3}$$

假设语言模型($P(S)$及$H(S)$)的参数不用优化，因此对于训练数据，(2)式中的互信息最大化就等价于(3)式的$H(S|X)$最小化。当训练数据中的样本及标记从独立同分布的分布提取，$H(S|X)$即为：

$$H(S|X)=-\frac{1}{R}\sum_{r=1}^{R}\log p(S_r|X_r,\Lambda)=-\frac{1}{R}\sum_{r=1}^{R}\log\frac{p(X_r,S_r|\Lambda)}{p(X_r)}.$$

因此，基于MMI判别学习的参数最优化可以通过最大化下面方程得到：

$$O_{MMI}(\Lambda)=\sum_{r=1}^{R}\log\frac{p(X_r,S_r|\Lambda)}{P(X_r)}=\sum_{r=1}^{R}\log\frac{p(X_r,S_r|\Lambda)}{\sum_{s_r}p(X_r,s_r|\Lambda)}, \tag{4}$$

其中，$P(s_r)$是模式序列s_r.的语言模型概率。

式(4)中的目标函数O_{MMI}是一个对数和形式，和接下来几个章节的判别训练准则相比，我们为式(4)构造了一个单调递增的幂函数。如下：

$$\widetilde{O}_{MMI}(\Lambda)=\exp[O_{MMI}(\Lambda)]=\prod_{r=1}^{R}\frac{p(X_r,S_r|\Lambda)}{\sum_{s_r}p(X_r,s_r|\Lambda)} \tag{5}$$

这里需要说明，$\widetilde{O}_{MMI}$和O_{MMI}有相同的最大值点的集合，因为最大值点对于单调递增函数是恒定不变的。为了和其它的判别训练准则相区别，记式(5)中的因子为：

$$\frac{p(X_r,S_r|\Lambda)}{\sum_{s_r}p(X_r,s_r|\Lambda)}=1-\sum_{s_r\neq S_r}P(s_r|X_r,\Lambda)$$

$$=1-\overbrace{\sum_{s_r}\underbrace{(1-\delta(s_r,S_r))}_{\text{0-1 loss}}P(s_r|X_r,\Lambda)}^{\text{model based expected loss}} \tag{6}$$

定义式(6)为标记X_r基于模型的期望效用(utility)，等于1减去基于模型的期望损失。

B. 最小分类误差(MCE)

基于MCE的分类器设计是模式识别中的一种基于判别函数(discriminant function)的方法[1,28,29]。分类器的判定准则被看做是判别函数集合的比较，当判定准则应用于分类器中时，参数估计包括使期望损失最小化。基于MCE判别学习的损失函数的构造，是以嵌在平滑函数形式中的分类器的识别误差率形式构造的，分类器的期望损失最小化直接关系到错分率的降低。

在基于MCE的判别学习中，目标(损失)函数可以利用基于似然度的生成模型通过以下步骤构建。对于每个训练标记X_r，判别函数的集合$\{g_s\}$为:

$$g_{s_r}(X_r;\Lambda) = \log p(X_r, s_r|\Lambda),$$

这里包含数据X_r和给定模型Λ下的模式序列(字符串)s_r的对数联合概率。分类或识别的判定规则定义为:

$$C(X_r) = s_r^* \qquad \text{iff } s_r^* = \underset{s_r}{\operatorname{argmax}}\, g_{s_r}(X_r;\Lambda).$$

实际上，MCE的判别学习中可考虑N个最容易混淆的竞争字符串，$s_{r,1}, \ldots, s_{r,N}$，与正确字符S_r间的竞争。这里最佳的N个字符串可以归纳定义为:

$$s_{r,1} = \underset{s_r: s_r \neq S_r}{\arg\max} \log p(X_r, s_r|\Lambda)$$

$$s_{r,i} = \underset{s_r: s_r \neq S_r, s_r \neq s_{r,1}, \ldots, s_{r,i-1}}{\arg\max} \log p(X_r, s_r|\Lambda), \quad i = 2, \ldots, N \tag{7}$$

其中Λ是分类器的当前参数模型集合。

错分率度量$d_r(X_r, \Lambda)$用来逼近对每个训练样本X_r判定准则的性能，当$d_r(X_r, \Lambda) \geq 0$表明有错分，$d_r(X_r, \Lambda) < 0$表明没有错分。实际上这样的错分类度量可以定义为:

$$d_r(X_r, \Lambda) = -g_{S_r}(X_r;\Lambda) + G_{S_r}(X_r;\Lambda) \tag{8}$$

其中$G_{S_r}(X_r;\Lambda)$是一个表示错误竞争字符串的分数的函数，$g_{S_r}(X_r;\Lambda)$是对正确字符串S_r.的判别函数。

对于只有一个最佳竞争字符串(one-best string)的MCE方法(N=1)，只有最易被混淆的错字符串s_{r1}才被认为是竞争对象，这里$G_{S_r}(X_r;\Lambda)$变为:

$$G_{S_r}(X_r;\Lambda) = g_{s_{r,1}}(X_r;\Lambda) \tag{9}$$

然而对于$N>1$的一般情况，$G_{S_r}(X_r;\Lambda)$可以使用多种定义。一个比较典型的定义形式为[29]:

$$G_{S_r}(X_r;\Lambda) = \log \left\{ \frac{1}{N} \sum_{i=1}^{N} p^{\eta}(X_r, s_{r,i}|\Lambda) \right\}^{\frac{1}{\eta}} \tag{10}$$

另外一种$g_{s_r}(X_r;\Lambda)$和$G_{S_r}(X_r;\Lambda)$典型的形式如下(这和式10很相似，见文献[54]):

$$\begin{cases} g_{S_r}(X_r;\Lambda) = \log p^{\eta}(X_r, S_r|\Lambda) \\ G_{S_r}(X_r;\Lambda) = \log \sum_{i=1}^{N} p^{\eta}(X_r, s_{r,i}|\Lambda) \end{cases} \tag{11}$$

其中，η为联合概率$p(X_r, s_r|\Lambda)$的尺度因子。在本文中，我们采用式(11)中的$G_{S_r}(X_r;\Lambda)$形式，同时为了计算方便设$\eta = 1$。($\eta \neq 1$的情况在[24]中进行讨论)。

给出错分率度量，对于每个训练样本r，损失函数可以通过Sigmoid函数来定义(见文献[28,29]):

$$l_r(d_r(X_r, \Lambda)) = \frac{1}{1 + e^{-\alpha d_r(X_r, \Lambda)}} \tag{12}$$

其中$\alpha > 0$表示sigmoid函数的斜率，通常靠经验决定。本文中为简单起见设定为1。在文献[25，p.156]有类似设定。这里需要说明式(12)的损失函数逼近记为平滑函数形式的0-1分类误差。

给定所有的模式序列集合$\{s_r\} = \{S_r, s_{r,1}, \ldots, s_{r,N}\}$和相应的观测数据$X_r$，其中$\eta = 1$，$\alpha = 1$，把(11)式代入(12)式，得到训练标记数据$X_r$的损失函数:

$$\begin{aligned} l_r(d_r(X_r, \Lambda)) &= \frac{\sum_{s_r, s_r \neq S_r} p(X_r, s_r|\Lambda)}{\sum_{s_r, s_r \neq S_r} p(X_r, s_r|\Lambda) + p(X_r, S_r|\Lambda)} \\ &= \frac{\sum_{s_r, s_r \neq S_r} p(X_r, s_r|\Lambda)}{\sum_{s_r} p(X_r, s_r|\Lambda)}. \end{aligned} \tag{13}$$

相应的，可定义效用函数为一减损失函数的形式，也就是:

$$u_r(d_r(X_r, \Lambda)) = 1 - l_r(d_r(X_r, \Lambda)). \tag{14}$$

基于MCE的判别学习的目标就变为对所有的训练数据而言，使训练期望损失最小化。

$$L_{MCE}(\Lambda)=\frac{1}{R}\sum_{r=1}^{R}l_r(d_r(X_r,\Lambda)). \tag{15}$$

很明显，使式(15)中的$L_{MCE}(\Lambda)$最小化等价于使下列MCE目标函数最大化。

$$O_{MCE}(\Lambda)=R(1-L_{MCE}(\Lambda))=\sum_{r=1}^{R}u_r(d_r(X_r,\Lambda))$$

$$=\sum_{r=1}^{R}\frac{p(X_r,S_r|\Lambda)}{\sum_{s_r}p(X_r,s_r|\Lambda)} \tag{16}$$

这里值得注意的是，MCE的求和形式(16)与MMI的求积形式(5)形成了强烈对比。

C. 最小音/词误差(MPE/MWE)

MPE/MWE是另一种判别学习方法，最早被文献[45,47]提出，并且在语音识别领域证明了其有效性。与MMI以及MCE不同，MMI/MCE是为适合大规模的模式序列（例如，字符串或者超字符串），MPE旨在提高子串模式水平的优化性能。在语音识别中，一组模式字符串通常和由一系列词组成的句子相对应，其中子字符串作为序列的组成部分可以是字或者音。

MPE需要最大化的目标函数定义为：

$$O_{MPE}(\Lambda)=\sum_{r=1}^{R}\frac{\sum_{s_r}p(X_r,s_r|\Lambda)A(s_r,S_r)}{\sum_{s_r}p(X_r,s_r|\Lambda)} \tag{17}$$

其中$A(s_r,S_r)$表示在句串s_r中的原始音(子串)精度（在文献[45,47]中提到）。原始音精度$A(s_r,S_r)$定义为参考字符串S_r的全部音的数目减去s_r的插入、删除和置换误差。

式(17)的MPE准则等价于对整个训练集原始音精度数目的基于模型的期望。这种关系可以将(17)式改写为：

$$O_{MPE}(\Lambda)=\sum_{r=1}^{R}\sum_{s_r}P(s_r|X_r,\Lambda)A(s_r,S_r)$$

其中$p(s_r|X_r,\Lambda)=\frac{p(X_r,s_r|\Lambda)}{p(X_r|\Lambda)}=\frac{p(X_r,s_r|\Lambda)}{\sum_{s_r}p(X_r,s_r|\Lambda)}$

是基于模型的后验概率。

式(17)中原始音精度$A(s_r,S_r)$可以广义的定义为原始子串精度。特别地，原始音精度$A_l(s_r,S_r)$亦能够以相同的方式定义为在参考字符串S_r的总字数减去s_r的插入、删除和置换误差。相似的，基于原始词精度$A_l(s_r,S_r)$,，可以得到MWE准则的等价定义：

$$O_{MWE}(\Lambda)=\sum_{r=1}^{R}\frac{\sum_{s_r}p(X_r,s_r|\Lambda)\,A_l(s_r,S_r)}{\sum_{s_r}p\ X_r,s_r|\Lambda} \tag{18}$$

因此本文将这两种方法均归为MPE/MWE一类。

D. 讨论

在单一训练样本级别时，MMI准则使用式(6)中基于模型的实用期望，此时MCE准则使用了由式(8)(12)和(14)定义的依赖分类器的平滑经验效用函数。MPE/MWE准则同样使用基于模型的期望效用函数，但是该效用是由子字符串计算得到的，例如，在音和字的层次中。本文为了计算方便，使用式(11)作为MCE的错分测度。因此，式(14)平滑的经验效用函数和式(6)有相同的形式。这可以直接用式(14)取代式(13)。

在多重训练样本级别时，通过比较式(5)、(16)、(17)和(18)，MMI训练使训练标记的模型预期效用乘积最大，此时MCE训练使对所有训练标记的平滑经验效用之和最大化，MPE/MWE训练则是使模型期望经验效用之和最大化。对效用函数(utility function)进行求和或乘积形式之间的差别，才是MMI和MCE/MPE/MWE之间的不同。这种不同造成了从MMI扩展原始GT/EBW方法到其它准则时的困难[47 p.92]。

接下来几章，本文将说明我们统一的学习准则如何反映这种不同性。

III. MMI、MCE、MPE/MWE目标函数的统一有理函数形式

本章给出了基于MMI、MCE和MPE/MWE准则的判别学习的目标函数，可以被映射到一个规范有理函数形式，并约束其分母函数为正数值。这种规范有理函数的形式有利于深入研究基于MMI，MCE和MPE/MWE的分类器。此外，这一统一的目标函数对分类器参数优化的统一框架发展起到了促进作用。

A. MMI目标函数的有理函数形式

基于式(5)的MMI目标函数的有理函数形式可以写为:

$$\widetilde{O}_{MMI}(\Lambda)=\frac{p(X_1\ldots X_R,S_1\ldots S_R|\Lambda)}{\sum_{s_1\ldots s_R}p(X_1\ldots X_R,s_1\ldots s_R|\Lambda)}=\frac{\sum_{s_1\ldots s_R}p(X_1\ldots X_R,s_1\ldots s_R|\Lambda)\,C_{MMI}(s_1\ldots s\ )}{\sum_{s_1\ldots s_R}p(X_1\ldots X_R,s_1\ldots s_R|\Lambda)} \tag{19}$$

其中,

$$C_{MMI}(s_1\ldots s_R)=\prod_{r=1}^{R}\delta(s_r,S_r) \tag{20}$$

是一个常数,该常数只与句序列$s_1,\ldots,s_R$相关,$\delta(s_r,S_r)$为克罗内尔δ函数,也就是:

$$\delta(s_r,S_r)=\begin{cases}1 & \text{if } s_r=S_r\\ 0 & \text{otherwise}\end{cases}.$$

对式(19),采用一般性假设,认为不同训练样本之间是相互独立的。

MMI目标函数旨在对整个训练数据而不是对每个单独的字符串进行条件似然函数的改进。它可以被看做是所有训练数据$s_1,\ldots,s_R$在超字符串层的判别性能测度,其中$C_{MMI}(s_1,\ldots,s_R)$可以看成超字符串$s_1,\ldots,s_R$的二元函数,当超字符串$s_1,\ldots,s_R$为正确时,取值为1,反之为0。

B. MCE目标函数的有理函数形式

和MMI的情况不同,式(16)给出了MCE目标函数,为若干个有理函数的总和而不是单独的有理函数。这导致应用GT的参数最优化框架去优化MCE的目标函数有困难。因此,MCE的目标函数通常使用广义概率下降算法(GPD)[9,28,29]或者其它基于梯度的方法[37,38]进行优化。尽管相当普及和很多成功应用,但基于GPD顺序学习的梯度下降主要存在两个缺点。第一,它是一种单样本循序学习算法。在计算时,对于GPD而言参数学习算法的并行化非常困难,而这对大规模任务很关键。第二,它不具有单调的学习算法,无法确定判别学习的终点。近年来基于梯度的批处理最优化方法的使用,包括批量-半批量的概率下降(batch and semi-batch probabilistic descent),快速传播(QuickProp),弹性反向传播(Rprop)算法,都对MCE算法进行了改进,并且提高了识别率[37,38]。然而,这些方法的单调收敛性还未确定。

本文使用一种不同的方法使MCE判别学习的目标函数适合于GT的参数优化。基于GT具有可扩展和单调收敛的学习性质,其备快速和稳定性优势。为了实现这种优势,需要重新对MCE函数进行规范,对MCE的目标函数提取一个规范且有理的函数形式。同时,MCE的规范有理的函数形式在该过程中有利于统一MCE与MMI、MPE/MWE的目标函数,这样 可以研究他们之间的异同。

MCE目标函数的有理函数形式可通过通分的方法推导如下:

$$O_{MCE}(\Lambda)=\sum_{r=1}^{R}\frac{\sum_{s_r}p(X_r,s_r|\Lambda)\delta(s_r,S_r)}{\sum_{s_r}p(X_r,s_r|\Lambda)} \tag{21}$$

$$=\underbrace{\frac{\sum_{s_1}p(X_1,s_1|\Lambda)\delta(s_1,S_1)}{\sum_{s_1}p(X_1,s_1|\Lambda)}}_{:=O_1}+\underbrace{\frac{\sum_{s_2}p(X_2,s_2|\Lambda)\delta(s_2,S_2)}{\sum_{s_2}p(X_2,s_2|\Lambda)}}_{:=O_2}$$

$$+\underbrace{\frac{\sum_{s_3}p(X_3,s_3|\Lambda)\delta(s_3,S_3)}{\sum_{s_3}p(X_3,s_3|\Lambda)}}_{:=O_3}+\cdots+\underbrace{\frac{\sum_{s_R}p(X_R,s_R|\Lambda)\delta(s_R,S_R)}{\sum_{s_R}p(X_R,s_R|\Lambda)}}_{:=O_R}$$

$$=\frac{\sum_{s_1}\sum_{s_2}p(X_1,s_1|\Lambda)p(X_2,s_2|\Lambda)[\delta(s_1,S_1)+\delta(s_2,S_2)]}{\sum_{s_1}\sum_{s_2}p(X_1,s_1|\Lambda)p(X_2,s_2|\Lambda)}$$

$$+O_3+\cdots+O_R$$

$$=\frac{\sum_{s_1s_2}p(X_1,X_2,s_1,s_2|\Lambda)[C_{MCE}(s_1s_2)]}{\sum_{s_1s_2}p(X_1,X_2,s_1,s_2|\Lambda)}+O_3+\cdots+O_R$$

$$= \frac{\sum_{s_1 s_2 s_3} p(X_1, X_2, X_3, s_1, s_2, s_3|\Lambda)[C_{MCE}(s_1 s_2 s_3)]}{\sum_{s_1 s_2 s_3} p(X_1, X_2, X_3, s_1, s_2, s_3|\Lambda)} + O_4 + \cdots + O_R$$

$$= \frac{\sum_{s_1 \ldots s_R} p(X_1 \ldots X_R, s_1 \ldots s_R|\Lambda)\, C_{MCE}(s_1 \ldots s_R)}{\sum_{s_1 \ldots s_R} p(X_1 \ldots X_R, s_1 \ldots s_R|\Lambda)} \quad (22)$$

其中 $C_{MMI}(s_1 \ldots s_R) = \sum_{r=1}^{R} \delta(s_r, S_r)$。$C_{MMI}(s_1, \ldots, s_R)$可以理解为$s_1, \ldots, s_R$,的字符串精度，值取为0到$R$之间的整数值，表示$s_1, \ldots, s_R$中正确的字符串的数量。MCE目标函数的有理函数形式(22)，在基于MCE的判别学习研究中将起到很关键的作用。

C. MPE/MWE目标函数的有理函数形式

与MCE类似，MPE/MWE的目标函数也是多个有理函数的总和，如在文章[47]中的讨论，很难直接推导出GT公式。为了避开这个问题，文献[45,47]中提出了一种弱性辅助函数(WSAF)的优化MPE/MWE目标函数的方法。而本文中，我们将MPE/MWE目标函数改为它的等价形式，即规范的有理函数形式，使得基于MPE/MWE的判别学习中的参数优化直接修正为基于GT的参数估计框架。这为统一的参数估计框架提供了可靠的单调收敛性质，而这种单调收敛性恰是如基于梯度和基于WSAF的逼近方法所缺少的。

我们发现，用在MCE目标函数的有理函数形式(22)中的推导方法，同样可以直接用来推导定义在式(17)和(18)中MPE/MWE目标函数的有理函数式。要注意的是，式(17)和(18)与式(21)的形式相同，除了$\delta(s_r, S_r)$是由$A(s_r, S_r)$或$A_l(s_r, S_r)$.取代。相同的MCE目标函数的推导步骤在这里同样可用，MPE/MWE的有理函数形式如下:

$$O_{MPE}(\Lambda) = \frac{\sum_{s_1 \ldots s_R} p(X_1 \ldots X_R, s_1 \ldots s_R|\Lambda)\, C_{MPE}(s_1 \ldots s_R)}{\sum_{s_1 \ldots s_R} p(X_1 \ldots X_R, s_1 \ldots s_R|\Lambda)} \quad (23)$$

这里 $C_{MPE}(s_1 \ldots s_R) = \sum_{r=1}^{R} A(s_r, S_r)$,

$$O_{MWE}(\Lambda) = \frac{\sum_{s_1 \ldots s_R} p(X_1 \ldots X_R, s_1 \ldots s_R|\Lambda)\, C_{MWE}(s_1 \ldots s_R)}{\sum_{s_1 \ldots s_R} p(X_1 \ldots X_R, s_1 \ldots s_R|\Lambda)} \quad (24)$$

$C_{MWE}(s_1 \ldots s_R) = \sum_{r=1}^{R} A_l(s_r, S_r)$.

$C_{MPE}(s_1, \ldots, s_R)$或$C_{MWE}(s_1, \ldots, s_R)$可被理解为超级字符串$s_1, \ldots, s_R$.中音或词的精确计数。它的上限值是所有训练数据（或者正确的超级字符串$S_1, \ldots, S_R$）中的音或词的总和数。但实际值有可能是负数，比如有太多的干扰错误时。相应地，$O_{MPE}(\Lambda)$和$O_{MWE}(\Lambda)$分别表示所有训练数据集的基于模型平均原声或词的精度计数。

D. 评价和讨论

这一部分主要讲这三个判别学习目标函数，MMI，MCE,和MPE/MWE，表示为一个统一的规范有理函数形式:

$$O(\Lambda) = \frac{\sum_{s_1 \ldots s_R} p(X_1 \ldots X_R, s_1 \ldots s_R|\Lambda) \cdot C_{DT}(s_1 \ldots s_R)}{\sum_{s_1 \ldots s_R} p(X_1 \ldots X_R, s_1 \ldots s_R|\Lambda)} \quad (25)$$

式(25)中$s = s_1 \ldots s_R$的总和表示所有R训练标记的不确定标记序列(包括正确的和不正确的)。因为还要进行进一步处理，所以在实际应用时不确定字符串数量会被大大减少。

在式(25)中，$X_1 \ldots X_R$表示在所有训练标记R中的全部观测数据序列（串）的集合，在把这些串连在一起形成单个串后，我们将 此单独的串称作超级字符串。$p_\Lambda(X_1 \ldots X_R, s_1 \ldots s_R)$是超级数据串$X_1 \ldots X_R$的联合概率分布，该超级数据串的不确定标记序列为$s_1 \ldots s_R$.。式(25)中，MMI，MCE和MPE/MWE由于依赖于准则的加权因子$C_{MMI}(s_1 \ldots s_R)$、$C_{MCE}(s_1 \ldots s_R)$和$C_{MPE}(s_1 \ldots s_R)$不同而相互区别。一个重要性质是：$C_{DT}(s_1 \ldots s_R)$仅由标记序列$s_1 \ldots s_R$,决定，与要进行最优化的参数集Λ无关。

用有理函数公式(25)做MMI,MCE和MPE/MWE的目标函数主要有两个目的。首先，MMI,MCE和MPE/MWE的目标函数被统一为一个规范的有理函数形式，因而能够研究不同的判别学习准则之间的关系，同时比较它们的性能。这能对不同的判别学习方法提供更好的观察。其次，式(25)表示的统一的目标函数克服了在判别学习中应用基于GT参数最优化框架的最主要障碍。这就为判别学习提供了一种可扩展的和普遍的参数评估框架，该框架具有高效且充分的算法收敛

特性。所有这些性能均是以前将判别学习应用于连续模式识别时的主要关注点。

如本节所提出的，MMI，MCE和MPE/MWE准则的有理函数形式之间的关键差别是式(25)分子中的加权因数。其中作为一个通用的加权因子，$C_{DT}(s_1 \ldots s_R)$取决于采用哪一种判别训练(DT)准则。例如，对于MMI，则有:

$$C_{DT}(s_1 \ldots s_R) = \prod_{r=1}^{R} \delta(s_r, S_r)$$

而对MPE，则有:

$$C_{DT}(s_1 \ldots s_R) = \sum_{r=1}^{R} A(s_r, S_r)$$

在MCE具有N个最佳竞争对手N>1的情况下:

$$C_{DT}(s_1 \ldots s_R) = \sum_{r=1}^{R} \delta(s_r, S_r)$$

而对于单个最优的MCE（即N=1），s_r仅属于子集$\{S_{r,}, S_{r,1}\}$。从式(25)的规则有理函数形式，可对MMI，MCE和MPE/MWE的目标函数进行直接对比。表1以表格形式给出了这些判别目标函数之间的关系。文献[47]指出，MPE/MWE与MCE和MMI有一个重要的区别，即对于错误字符 串的MPE/MWE标准的加权给定值，取决于错误字符串中子字符串的数目。根据整个句串是否正确，MCE和MMI产生了一个二进制的区别，或许不适用于以减少子字符串错误为目的的情况。对于MCE，该区别可以通过比较二值函数和明确得出:

$$C_{DT}(s_1 \ldots s_R) = \sum_{r=1}^{R} \delta(s_r, S_r)$$

对于MPE/MWE，非二值函数的和为:

$$C_{DT}(s_1 \ldots s_R) = \sum_{r=1}^{R} A(s_r, S_r)$$

该关键差别造成了MPE/MWE和MCE中的子字符串级别及字符串级别的识别性能优化之间的不同。进而，MMI采用二进制函数的乘积形式代替了MCE中的二进制函数的和形式:

$$C_{DT}(s_1 \ldots s_R) = \prod_{r=1}^{R} \delta(s_r, S_r)$$

由上式可得，MMI在超字符串级别上得到了性能优化，举例来说，如果任一句的标记是错误的，克罗内克三角函数的联合乘积即变为零。因此，除了与一个正确标签/抄本的序列相对应，式(25)分子上的所有条件和都为0。正如在语音判别实验[35]，[45]-[47]中被广泛观测的那样，MMI标准不如MCE或MPE/MWE令人满意。

表一 $C_{DT}(S_1 \ldots S_R)$ IN THE UNIFIED RATIONAL-FUNCTION FORM FOR MMI, MCE, AND MPE/MWE OBJECTIVE FUNCTIONS. THE SET OF "COMPETING TOKEN CANDIDATES" DISTINGUISHES *N*-BEST AND ONE-BEST VERSIONS OF THE MCE. NOTE THAT THE OVERALL $C_{DT}(S_1 \ldots S_R)$ IS CONSTRUCTED FROM ITS CONSTITUENTS $C_{DT}(S_R)$'S IN INDIVIDUAL STRING TOKENS BY EITHER SUMMATION (FOR MCE, MPE/MWE) OR PRODUCT (FOR MMI).

目标函数	$C_{DT}(s_r)$	$C_{DT}(s_1 \ldots s_R)$	DT中所用的标记序列集合
MCE (N-BEST)	$\delta(s_r, S_r)$	$\sum_{r=1}^{R} C_{DT}(s_r)$	$\{S_r, S_{r,1}, \ldots, S_{r,N}\}$
MCE (ONE-BEST)	$\delta(s_r, S_r)$	$\sum_{r=1}^{R} C_{DT}(s_r)$	$\{S_r, S_{r,1}\}$
MPE	$A(s_r, S_r)$	$\sum_{r=1}^{R} C_{DT}(s_r)$	所有可能的标记序列
MWE	$A_l(s_r, S_r)$	$\sum_{r=1}^{R} C_{DT}(s_r)$	所有可能的标记序列
MMI	$\delta(s_r, S_r)$	$\prod_{r=1}^{R} C_{DT}(s_r)$	所有可能的标记序列

从式(25)目标函数的统一形式中得到的另一个结论是，在训练数据只有一个样本标记(R=1)的特例中，如假设这时该句只包含一个音，那么这三种标准（即MMI,MCE和MPE/MWE标准）即为相同。该结论十分明显，因为在这种情况下，$C_{DT}(s_1 \ldots s_R)$完全相同。只有当训练集合包含多重语句标记的情况时，它们之间的不同才会出现。在多个训练样本情况下，当有理函数形式(25)对于三种规则均保持不变时，差别主要在独立于Λ的加权因子$C_{DT}(s_1 \ldots s_R)$中体现。

在序列模式识别中，尽管我们尝试对MMI，MCE和MPE/MWE的三种目标函数形式推导出基于GT的参数最优化框架，但应该注意的是该统一的目标函数(25)能为推导出其它参数最优化方法提供一个关键性基础。例如，最近Jebara在文献[26][27]提出了一种有理函数的参数最优化方法可作为GT方法一种替代。该方法是基于反转的Jensen不等式，基于此，一种对带有指数族密度HMMs的最佳解决方案能得以构建。

IV. 采用GT的有理函数优化

基于GT的参数最优化是指一种批处理方式的，迭代的，最优化方案。目标函数的值随每次迭代而增加。也就是说，新的模型参数集Λ通过$\Lambda = T(\Lambda')$变换利用当前的模型Λ'估计得出，其性质是目标函数值会不停增

加$O(\Lambda) > O(\Lambda')$，除非$\Lambda = \Lambda'$.。估计HMM参数时，EBW算法是这种优化技术类型的一种典型。GT/EBW算法最初是由Baum和他的同事因为齐次多项式提出的[3],[4]。它后来被扩展到用来优化非齐次的有理函数，如文献[18]中所载。EBW算法开始流行是因为在离散HMMs的MMI判别训练中的成功应用[18]。它后来也被扩展及应用到连续密度HMMs的MMI判别训练中[2],[20],[41],[59],[61]。

GT/EBW算法的重要性在于它的单调收敛性、算法的高效性、并行执行时的可扩展性以及用于大规模最优化问题时的的解析解的参数更新公式。GT的统一参数优化框架还减轻了对其它经验设置的需求，例如，在其它方法中，调整参数由经验设置的学习速率决定[29][52]。

令$G(\Lambda)$和$H(\Lambda)$为参数集Λ的两个实值函数，且分母上那个的函数$H(\Lambda)$为正值。基于GT的参数优化目标就是找到一个最优的Λ使得目标函数$O(\Lambda)$最大，而这个目标函数是一个如下表示的有理函数:

$$O(\Lambda) = \frac{G(\Lambda)}{H(\Lambda)}. \tag{26}$$

举例说明，$O(\Lambda)$可以成为式(19)，(22)，(23)和(24)所示的有理函数中的一个，这四个方程分别为MMI，MCE和MPE/MWE的目标函数，或者为广义有理函数式(25)。对于广义情况下的式(25)，我们有:

$$G(\Lambda) = \sum_s p(X, s|\Lambda)\, C(s), \text{ and } H(\Lambda) = \sum_s p(X,\ s|\Lambda) \tag{27}$$

其中，我们使用速记符号$s = s_1 \ldots s_R$来表示所有R训练标记/句子的被标序列，并用$X = X_1 \ldots X_R$,来表示对所有R训练标记的观测数据序列。

首要辅助函数

正如在文献[18]中最初提出的，对于目标函数式(26)，基于GT的优化算法将构建一个如下形式的辅助函数:

$$F(\Lambda;\Lambda') = G(\Lambda) - O(\Lambda')H(\Lambda) + D \tag{28}$$

其中D是一个与参数集合相独立的量，通过将GT应用于一个已存在的模型参数集Λ'从而估计新的模型参数集合Λ。这种GT算法开始于（初始）参数集合Λ'（例如，利用最大似然度(ML)训练获得）。然后，通过最大化辅助函数F $F(\Lambda;\Lambda')$,该算法将参数集合从Λ'更新为Λ，这个迭代过程止于达到收敛条件。使该辅助函数$F(\Lambda;\Lambda')$最大化通常能比使原始有理函数$O(\Lambda)$的最大化更为可行。基于GT的参数优化的重要性质就是，只要D是与参数集Λ无关的量，$F(\Lambda;\Lambda')$的增加就会确保$O(\Lambda)$的增加。这从下面的推导可容易看出:

用$\Lambda = \Lambda'$代如式(28)中，可得:

$$\underbrace{G(\Lambda') - O(\Lambda')\, H(\Lambda')}_{=0} + D = D.$$

因此，

$$F(\Lambda;\Lambda') - F(\Lambda';\Lambda') = F(\Lambda;\Lambda') - D = G(\Lambda) - O(\Lambda')H(\Lambda)$$
$$= H(\Lambda)\left(\frac{G(\Lambda)}{H(\Lambda)} - O(\Lambda')\right) = H(\Lambda)(O(\Lambda) - O(\Lambda'))$$

因为$H(\Lambda)$是正数，所以在左边$O(\Lambda) - O(\Lambda')$时，即有右边的$F(\Lambda;\Lambda') - F(\Lambda';\Lambda')$。

次要辅助函数

$F(\Lambda;\Lambda')$直接优化的时候依旧可能比较困难，因此一种次要辅助函数可以通过前面的辅助函数$F(\Lambda;\Lambda')$.进行构造和优化。如文献[17]提到的那样，在基于GT的参数估计中次要辅助函数具有如下结构:

$$V(\Lambda;\Lambda') = \sum_s \sum_q \sum_\chi f(\chi, q, s, \Lambda') \log f(\chi, q, s, \Lambda) \tag{29}$$

$f(\chi, q, s, \Lambda)$是一个由离散变量χ, q, s构建的正值函数，与前面提到的主要辅助函数相关。

$$F(\Lambda;\Lambda') = \sum_s \sum_q \sum_\chi f(\chi, q, s, \Lambda) \tag{30}$$

通过将Jensen不等式应用到凸的对数函数中，容易证明出函数$V(\Lambda;\Lambda')$的增加能够保证函数$\log F(\Lambda;\Lambda')$.的增加。因为对数是单调增加的函数，所以这意味着函数$F(\Lambda;\Lambda')$的增加，因此原始的目标函数$O(\Lambda)$也会增加。

V. 基于GT框架的离散HMM中的判别学习

对离散HMMs中的基于GT-EBW的判别学习需要估计参数模型$\Lambda = \{\{a_{i,j}\}, \{b_i(k)\}\}$，包括状态转移概率和发射概率。我们推导出参数优化公式能够使式(25)及其中覆盖的MMI，MCE和MPE/MWE的判别目标函数$O(\Lambda)$产生增长。判别函数$O(\Lambda)$很难直接优化。尽管它是一个有理函数。但应用服从基于GT/EBW的参数估计框架，我们可以先构建辅助函数F，然后基于F构建

次要辅助函数V。我们将会阐述如何优化$V(\Lambda;\Lambda')$，导出三种判别准则MMI，MCE和MPE/MWE的参数估计公式。只要目标函数能用式(25)中的有理函数形式表示，该方法就可以应用在其它任何判别准则中。

对于离散隐马尔科夫模型(DHMM)，观测空间可以用一些离散码本进行量化。因此，$X = X_1 \dots X_R$，是所有训练标记的串联，每一个训练标记X_r包括一列离散索引，这些指针通过把第r个标记的观测时间序列映射到每个离散索引序列的元素$x_{r,t} \in [1,2,\dots,K]$得到，其中，$K$是码本索引集合的取值范围，$x_{r,t}$是索引组成部分，用来量化第$r$个标记索引处第$t$帧的观测数据。

A. 构造首要辅助函数

把式(27)代入式(28)。我们可以得到下面的辅助函数

$$\begin{aligned} F(\Lambda;\Lambda') &= \sum_s p(X,s|\Lambda)C(s) - O(\Lambda')\sum_s p(X,s|\Lambda) + D \\ &= \sum_s p(X,s|\Lambda)[C(s) - O(\Lambda')] + D \\ &= \sum_s \sum_q p(X,q,s|\Lambda)[C(s) - O(\Lambda')] + D \end{aligned} \tag{31}$$

这里q是一个隐马尔科夫模型(HMM)状态序列，$s = s_1 \dots s_R$是所有R训练标记的超级标记序列(包括正确的和错误的句子)。上面的辅助函数$F(\Lambda;\Lambda')$的主要条件可以理解为对准确值的期望偏差。

B. 构建次要辅助函数$(\Lambda;\Lambda')$

因为$p(s)$取决于语言模型且与优化Λ不相关，因此，我们可以得到$p(X,q,s|\Lambda) = p(s)\cdot p(X,q|s,\Lambda)$，以及：

$$\begin{aligned} F(\Lambda;\Lambda') &= \sum_s \sum_q [C(s) - O(\Lambda')]p(s)p(X,q|s,\Lambda) + D \\ &= \sum_s \sum_q \sum_\chi [\Gamma(\Lambda') + d(s)]p(\chi,q|s,\Lambda) \end{aligned} \tag{32}$$

这里

$$\begin{aligned} \Gamma(\Lambda') &= \delta(\chi,X)p(s)[C(s) - O(\Lambda')] \\ D &= \sum_s d(s) \end{aligned} \tag{33}$$

D是一个独立于参数集Λ的值。在式(33)中，$\delta(\chi,X)$是克罗内克尔δ函数，χ表示χ所属的整个离散数据空间。采用文献[20]的思想，这里该数据空间总和的引入是为了满足式(28)和(32)中常数D成为独立参数的条件。也就是说，在(32)中：

$$\sum_s \sum_q \sum_\chi d(s)p(\chi,q|s,\Lambda) = \sum_s d(s) = D$$

是一个独立于Λ的常数。尽管此和函数独立于Λ，为了补偿$\Gamma(\Lambda')p(\chi,q|s,\Lambda)$为负的可能性，每一项$d(s)p(\chi,q|s,\Lambda)$的是$\Lambda$相关的。下面进行详细叙述。

为了在式(32)的函数基础上构建式(29)的次要辅助函数，我们首先根据式(30)从式(32)可得：

$$f(\chi,q,s,\Lambda) = [\Gamma(\Lambda') + d(s)]p(\chi,q|s,\Lambda)$$

为了保证上面的$f(\chi,q,s,\Lambda)$是正的，$d(s)$应该选一个充分大的值这样才能保证式(32)中的$\Gamma(\Lambda') + d(s) > 0$。然后根据式(29)，我们可以得出：

$$\begin{aligned} V(\Lambda;\Lambda') &= \sum_q \sum_s \sum_\chi [\Gamma(\Lambda') + d(s)]p(\chi,q|s,\Lambda') \\ &\quad \times \log\Big\{\underbrace{[\Gamma(\Lambda') + d(s)]}_{\text{optimization-independent}}p(\chi,q|s,\Lambda)\Big\} \\ &= \sum_q \sum_s \sum_\chi [\Gamma(\Lambda') + d(s)]p(\chi,q|s,\Lambda')\log p(\chi,q|s,\Lambda) \\ &\quad + \text{Const.} \\ &= \sum_q \sum_s p(X,q,s|\Lambda')(C(s) - O(\Lambda'))\log p(X,q|s,\Lambda) \\ &\quad + \sum_q \sum_s \sum_\chi d(s)p(\chi,q|s,\Lambda')\log p(\chi,q|s,\Lambda) \\ &\quad + \text{Const.} \end{aligned} \tag{34}$$

辅助函数(34)比函数(32)更容易优化，因为在函数(34)中所用到的新对数$\log p(X,q|s,\Lambda)$(函数(32)中就没有用到)可以对$V(\Lambda;\Lambda')$进行更加有效的简化，我们接下来会进行描述。

C. 简化次要辅助函数$(\Lambda;\Lambda')$

首先我们先不管函数(34)中所用到的独立优化的常数，为了将联合概率$p(X,q,s|\Lambda')$转化为后验概率$p(q,s|X,\Lambda')=p(s|X,\Lambda')p(q|X,s,\Lambda')$.，用另外一个独立优化值来分割$V(\Lambda;\Lambda')$。我们可以得到一个等价的辅助函数：

$$\begin{aligned} U(\Lambda;\Lambda') &= \sum_q \sum_s p(s|X,\Lambda')p(q|X,s,\Lambda')(C(s) - O(\Lambda')) \\ &\quad \times \log p(X,q|s,\Lambda) \\ &\quad + \sum_q \sum_s \sum_\chi d'(s)p(\chi,q|s,\Lambda')\log p(\chi,q|s,\Lambda) \end{aligned} \tag{35}$$

这里

$$d'(s) = d(s)/p(X|\Lambda'). \tag{36}$$

因为X仅仅取决于隐马尔科夫模型(HMM)的状态序列q，因此我们有$p(X,q|s,\Lambda)=p(q|s,\Lambda)\cdot p(X|q,\Lambda).$。进而$U(\Lambda;\Lambda')$可以进一步被分解为以下四个部分：

$$\begin{aligned}U(\Lambda;\Lambda')=&\sum_q\sum_s\overbrace{p(s|X,\Lambda')p(q|X,s,\Lambda')(C(s)-O(\Lambda'))}^{term\text{-}I}\log p(X|q,\Lambda)\\&+\sum_q\sum_s\sum_\chi\underbrace{d'(s)p(\chi,q|s,\Lambda')}_{term\text{-}II}\log p(\chi|q,\Lambda)\\&+\sum_q\sum_s\overbrace{p(s|X,\Lambda')p(q|X,s,\Lambda')(C(s)-O(\Lambda'))}^{term\text{-}III}\log p(q|s,\Lambda)\\&+\sum_q\sum_s\sum_\chi\underbrace{d'(s)p(\chi,q|s,\Lambda')}_{term\text{-}IV}\log p(q|s,\Lambda)\end{aligned}\tag{37}$$

在这里，$X=X_1\ldots X_R$,集合了所有R独立句子标记的训练数据。对于每一个标记$X_r=x_{r,1},\ldots,x_{r,Tr}$,观测向量$x_{r,t}$是相互独立的，且它只取决于隐马尔科夫模型(HMM)在时间T时的状态。因此，$\log p(X|q,\Lambda)$是可以被分解的，从而能对式(37)中的第一项和第二项进行简化。为了简化(37)中的第三项和第四项，我们根据的一阶隐马尔科夫模型HMM的特性——时间t时的状态仅依赖于时间t-1时的状态，从而对$\log p(q|s,\Lambda)$进行分解。我们可以得到分解和优化后的目标函数：

$$U(\Lambda;\Lambda')=U_1(\Lambda;\Lambda')+U_2(\Lambda;\Lambda')\tag{38}$$

这里$U_1(\Lambda;\Lambda')$表示式(37)中第一项和第二项的和，它只是和发射概率$b_i(k)$的优化有关，$U_2(\Lambda;\Lambda')$市第三项和第四项的和，仅和来优化转移概率$a_{i,j}$.相关。

D. 辅助函数优化中的增长转型估计

为了优化在限制条件$\sum_{k=1}^K b_i(k)=1$下的转移概率$\sum_{k=1}^K b_i(k)=p(x_{r,t}=k|q_{r,t}=i,\Lambda)$，$k=1,2,\ldots,K,$，我们使用拉格朗日乘法构建下面的函数：

$$W_1(\Lambda;\Lambda')=U_1(\Lambda;\Lambda')+\sum_{i=1}^I\lambda_i\left(\sum_{k=1}^K b_i(k)-1\right)\tag{39}$$

设置值为$\frac{\partial W_1(\Lambda;\Lambda')}{\partial\lambda_i}=0$和$\frac{\partial W_1(\Lambda;\Lambda')}{\partial b_i(k)}=0,\ k=1,\ldots,K$，得到如下$K$+1个等式：

$$\begin{aligned}&\sum_{k=1}^K b_i(k)-1=0\\0=\lambda_i b_i(k)+&\sum_{r=1}^R\sum_{\substack{t=1\\s.t.x_{r,t}=k}}^{T_r}\overbrace{\sum_s p(s|X,\Lambda')(C(s)-O(\Lambda'))\gamma_{i,r,s_r}(t)}^{\Delta\gamma(i,r,t)}\end{aligned}$$

$$+\sum_{r=1}^R\sum_{t=1}^{T_r}d(r,t,i)b_i'(k),\ k=1,\ldots,K.$$

这里两边均乘以$b_i(k)$。为了求解$b_i(k)$，我们定义：

$$D_i=\sum_{r=1}^R\sum_{t=1}^{T_r}d(r,t,i)\tag{40}$$

$$\Delta\gamma(i,r,t)=\sum_s p(s|X,\Lambda')(C(s)-O(\Lambda'))\gamma_{i,r,s_r}(t)\tag{41}$$

这里$d(r,t,i)=\sum_s d'(s)p(q_{r,t}=i|s,\Lambda')$,

我们可以得到重估公式：

$$b_i(k)=\frac{\sum_{r=1}^R\sum_{\substack{t=1\\s.t.x_{r,t}=k}}^{T_r}\Delta\gamma(i,r,t)+b_i'(k)D_i}{\sum_{r=1}^R\sum_{t=1}^{T_r}\Delta\gamma(i,r,t)+D_i}\tag{42}$$

现在我们定义：

$$\tilde{D}_i=\sum_{r=1}^R\sum_{t=1}^{T_r}d(r,t-1,i)\tag{43}$$

$$\Delta\xi(i,j,r,t)=\sum_s p(s|X,\Lambda')(C(s)-O(\Lambda'))\xi_{i,j,r,s_r}(t)\tag{44}$$

可得：

$$a_{i,j}=\frac{\sum_{r=1}^R\sum_{t=1}^{T_r}\Delta\xi(i,j,r,t)+a_{i,j}'\tilde{D}_i}{\sum_{r=1}^R\sum_{t=1}^{T_r}\Delta\gamma(i,r,t)+\tilde{D}_i}\tag{45}$$

参量重估公式(42)和(45)是MMI，MCE和MPE/MWE的统一重估公式。对此统一的重估公式,MMI，MCE和MPE/MWE最主要的不同就在于，式(41)中$\Delta\gamma(i,r,t)$和式(44)中$\Delta\xi(i,j,r,t)$的权重项导致的约束$C(s)$不同计算。具体的MMI，MCE和MPE/MWE计算$\Delta\gamma(i,r,t)$的细节详见附件I和文献[24]。

E. 设置常量D_i

式(43)中常量D_i和式(60)中$\tilde{D}_i$值决定了上述GT/EBW算法的稳定性和收敛速度。设置D_i的理论基础是需要式(32)中描述的式(58)$d(s)$足够大，从而使$\Gamma(\Lambda')+d(s)>0$.。从式(33)中可得：

$$\begin{aligned}\Gamma(\Lambda')&=\delta(\chi,X)p(s)[C(s)-O(\Lambda')]\\&=\begin{cases}p(s)[C(s)-O(\Lambda')] & \text{if } \chi=X\\0\end{cases}\end{aligned}$$

因此，$d(s) > \max\{0, -p(s)[C(s) - O(\Lambda')]\}$。给出：

$$D_i > \frac{1}{p(X|\Lambda')} \sum_{r=1}^{R} \sum_{t=1}^{T_r} \sum_{s} \max\{0,\ p(s)[O(\Lambda') - C(s)]\} \\ p(q_{r,t} = i|s, \Lambda') \tag{46}$$

类似的，我们可以推导出：

$$\widetilde{D}_i > \frac{1}{p(X|\Lambda')} \sum_{r=1}^{R} \sum_{t=1}^{T_r} \sum_{s} \max\{0,\ p(s)[O(\Lambda') - C(s)]\} \\ p(q_{r,t-1} = i|s, \Lambda'). \tag{47}$$

实际上，式(56)和(60)中给出的Di和$\widetilde{D}_i$过于保守且值过大，这会导致比以前的经验方法更慢的收敛速度。在这篇回顾中我们不会讨论这类具体的启示，但仍指出这依然是一个感兴趣的研究问题，且会向读者推荐这类研究或者讨论，见文献[18],[41],[42],[47],[54],[59]和[61]。

VI. 对连续密度HMMs下的区别性学习

对连续密度HMMs (CDHMMs)来说，观测空间并没有被量化。由于这个原因，$X = X_1 \ldots X_R$,是所有训练标记的串联，每一个训练标记Xr包括一系列的连续随机变量。我们将公式(25)应用于CDHMMS的判别学习中。特别地，在之前的等式(29)和(30)中χ.是一个连续变量，因此χ.域内的总和均被改为χ.域的积分，这样

$$V(\Lambda;\Lambda') = \sum_{s} \sum_{q} \int_{\chi} f(\chi,q,s,\Lambda') \log f(\chi,q,s,\Lambda) d\chi \tag{48}$$

这里积分函数$f(\chi,q,s,\Lambda)$被定义为：

$$F(\Lambda;\Lambda') = \sum_{s} \sum_{q} \int_{\chi} f(\chi, q, s, \Lambda) d\chi \tag{49}$$

相应地，

$$F(\Lambda;\Lambda') = \sum_{s} \sum_{q} [C(s) - O(\Lambda')] p(s) p(X, q|s,\Lambda) + D \\ = \sum_{s} \sum_{q} \int_{\chi} [\Gamma(\Lambda') + d(s)] p(\chi, q|s, \Lambda) d\chi \tag{50}$$

这里，

$$f(\chi,q,s,\Lambda) = [\Gamma(\Lambda') + d(s)] p(\chi,q|s,\Lambda) \tag{51}$$

$$\Gamma(\Lambda') = \delta(\chi, X) p(s) [C(s) - O(\Lambda')] \tag{52}$$

式(52)中的$\delta(\chi, X)$是Dirac函数。

在经过与前面部分类 似的推导后，式(45)中的转移概率估计公式同样存在于离散隐马尔科夫模型HMM中。

A. 基于GT的参数估计在高斯密度CDHMM中的应用

我们首先推导CDHMM中的具有高斯分布的GT参数估计公式，然后再将其推广到后面混合高斯分布的情况。对于高斯分布的CDHMM，观测概率密度函数$b_i(x_{r,t})$以如下形式为高斯分布：

$$b_i(x_{r,t}) \propto \frac{1}{|\Sigma_i|^{1/2}} \exp\left[-\frac{1}{2}(x_{r,t} - \mu_i)^T \Sigma_i^{-1} (x_{r,t} - \mu_i)\right]. \tag{53}$$

这里$(\mu_i, \Sigma_i), i = 1, 2, \ldots, I$是状态$i$时高斯分布的平均向量和协方差矩阵。

为了求出μ_i和Σ_i，我们设定$\frac{\partial U_1(\Lambda;\Lambda')}{\partial \mu_i} = 0; \frac{\partial U_1(\Lambda;\Lambda')}{\partial \Sigma_i} = 0.$:

通过求解μ和Σ_i给出GT的公式，其中$\Delta\gamma(i,r,t)$如式(41)中定义，D_i如(40)中定义。

$$\mu_i = \frac{\sum_{r=1}^{R} \sum_{t=1}^{T_r} \Delta\gamma(i,r,t) x_t + D_i \mu_i'}{\sum_{r=1}^{R} \sum_{t=1}^{T_r} \Delta\gamma(i,r,t) + D_i}, \tag{54}$$

$$\Sigma_i = \frac{\sum_{r=1}^{R} \sum_{t=1}^{T_r} [\Delta\gamma(i,r,t)(x_t - \mu_i)(x_t - \mu_i)^T] + D_i \Sigma_i' + D_i(\mu_i - \mu_i')(\mu_i - \mu_i')^T}{\sum_{r=1}^{R} \sum_{t=1}^{T_r} \Delta\gamma(i,r,t) + D_i} \tag{55}$$

正如在离散隐马尔可夫模型的例子中那样，式(54)和(55)是基于式(25)中广义判别目标函数$O(\Lambda)$的形式，其中涵盖MMI，MCE和MPE/MWE等特例。这致使统一的基于GT参数估值的MMI，MCE和MPE/MWE方程与其它任何判别目标函数一样，都可以映射为式(25)的有理函数形式。此外，式(54)和(55)中$\Delta\gamma(i,r,t)$的定义为与离散隐马尔科夫模型例子中式(41)相同，对于MMI，MCE和MPE/MWE，唯一的区别就是$C(s)$，而这将在附件和文献[24]中做进一步说明。

B. CDHMM中常量D_i的设置

根据Jensen不等式，设定一个适合 的常量D_i来确保式(54)和(55)为增长转型的理论基础是式(32)指出的要求，即式(58)中的$d(s)$需要足够大以确保对于任何字串s和任何观测序列χ，满足$\Gamma(\Lambda')+d(s)>0$，其中$\Gamma(\Lambda')=\delta(\chi,X)p(s)[C(s)-O(\Lambda')]$在式(33)中定义。但是，CDHMM中的$\delta(\chi,X)$是狄拉克(Dirac)三角函数，它的分布满足中心点时具有密度函数的极大值，也就是说，当$\chi=X$时$\delta(\chi,X)=+\infty$。所以，对于字串s，使$C(s)-O(\Lambda')<0,\Gamma(\Lambda')|_{\chi=X}=-\infty$.。在这种条件下，要找到一个边界$d(s)$来满足$\Gamma(\Lambda')+d(s)>0$是不可能的，所以不能采用Jensen不等式。请注意，这个问题不会出现在离散隐马尔科夫模型中，因为那时$\delta(\chi,X)$是Kronecker函数，只能取有限值0或1。

如果能证明出存在足够大且有界的常量D使式(48)中的$V(\Lambda;\Lambda')$可能由式(51)定义为积分形式，那就仍是有效得辅助函数$F(\Lambda;\Lambda')$；也就是说$V(\Lambda;\Lambda')$值的增加能够保证$F(\Lambda;\Lambda')$值的增 加，那么上面提到的CDHMMs难题就可以克服，且可以继续采用同样的推导方式。关于CDHMMs中基于GT的MMI训练近年来由文献[2]发展和研究，它同时也支持本文的广义有理函数判别训练准则。因此有界的边界D_i是存在的（技术细节见文献[24]）。

尽管一个足够大的D_i值能确保GT迭代估计公式的单调收敛性，也就是说，对于离散HMM采用式(42)和(45)，对于CDHMM采用式(54)和(55)，通过单调收敛性证明得到的D_i值是一个宽松的不确定上限，对合理收敛速度而言它可能过大。事实上，我们通常经验地设置D_i值从而得到折衷的训练性能。更多探讨和对比详见文献[23][41][42][45][46][47][54][59][60][61][66]。

VII. 相关工作和讨论

A. 与其它方法的关系

在最近的文献[35]和[54]中提出一种能够统一多种判别学习的方法 ，包括MMI，MCE和MPE/MWE，（早些的文章[54]中未包括MPE/MWE）。文献[35],[54]探讨了MMI，MCE和MPE/MWE准则中目标函数的异同点。本文提出的框架需要一个额外步骤来统一这些准则为正则有理函数形式(25)，基于GT的判别学习应用于该有理函数中，包括作为特例的MMI，MCE和MPE/MWE标准。这从两个方面来看都是重要的：首先，在基于统一有理函数形式(25)的目标函数中，提供了对MMI，MCE和MPE/MWE准则之间本质关系的更准确和更直接的见解。其次，使一个统一的基于GT的参数优化框架能应用于MMI，MCE和MPE/MWE和其他判别标准中去。

在文献[35]中提到的框架是基于下面的目标函数形式（用本文中的数学符号重写了上述公式，以利于比较）：

$$O(\Lambda)=\frac{1}{R}\sum_{r=1}^{R}f\left(\frac{1}{\eta}\log\frac{\sum_{s_r}p^{\eta}(X_r,s_r|\Lambda)C_{DT}(s_r)}{\sum_{s_r\in M_r}p^{\eta}(X_r,s_r|\Lambda)}\right)\qquad(56)$$

其中$C_{DT}(s_r)$与表1取值相同。式(56)中平滑函数$f(z)$，竞争词序列M_r和的权值 的选择是由表2提供，专为不同类型DT准则而建立。在表2中q是平滑函数的斜率。

方程(56)表明不同的判别准则具有一个相似的核函数形式，且因依赖准则且能调整核的平滑函数$f(z)$而所有不同，其中目标函数是平滑函数的和。方程(56)是对MMI，MCE，MPE/MWE目标函数的一般性描述。然而由于非线性函数$f(z)$的存在，它并不是有理函

数（定义为两个多项式函数的比）的一般形式。对于这些准则间求和与乘积之间最大的区别没有明确提出。在本文提出的方法中，我们直接提出了该问题，并揭示了MMI，MCE和MPE/MWE准则的目标函数具有一个确定的有理函数形式(25)，且对于每个判别标准，目标函数只会因独立于模型的量值$C_{DT}(s_1 \ldots s_R)$.而出现不同。

表二 CHOICES OF THE SMOOTHING FUNCTION *f(z)*, ALTERNATIVE WORD SEQUENCES M_r, AND EXPONENT WEIGHT η IN (56) FOR VARIOUS TYPES OF DT CRITERIA. THIS IS MODIFIED FROM THE ORIGINAL TABLE IN [54].

准则	光滑函数*f(z)*	词序选择 M_R	η
MCE (*N*-BEST)	$-1/[1+\exp(2qz)]$	$\{s_r\}$ EXCLUDING S_r	≥1
MCE (ONE-BEST)	$-1/[1+\exp(2qz)]$	$\{s_{r,1}\}$	N/A
MPE/MWE	$\exp(z)$	ALL POSSIBLE LABEL SEQUENCE $\{s_r\}$	1
MMI	z	ALL POSSIBLE LABEL SEQUENCE $\{s_r\}$	1

如表2所示，由于*f(z)*对MPE/MWE和MCE准则而言是一个非线性函数，原始GT的解[18]是并不能直接应用于MPE/MWE和MCE准则的目标函数的（因为当*f(z)*为非线性时试(56)不满足GT的要求）。为了避开该难题，文献[30]进行了研究。在[30]中，原始目标函数是逼近一个泰勒级数的展开式。然后通过和文献[18]中相似的方法，基于GT的参数优化可以被用于泰勒级数展开的部分和，它是一个有限级数的多项式。这形成了早期的基于GT的MCE和MPE/MWE理论基础[35][54]。然而，部分和的正值增长是依赖于部分和级数的（在文献[18]中可见关于这点更多讨论），而且当等级接近无限大时它将消失。它甚至消失的比泰勒级数逼近的误差还快。因此它并没有明确表明重估公式能确保目标函数值以迭代方式增长。

相反地，本文中描述的统一有理函数方法与文献[35]和[54]所做的工作不一样。它不依赖于泰勒级数逼近，且表明了MMI，MCE和MPE/MWE准则的目标函数具有一个明确的广义有理函数形式(25)。因此基于GT的参数优化框架能以一种建设性的方法直接应用于式(25)。在本文中采取的方法是在文献[2]和[20]工作基础上所做，比文献[3]和[18]的工作更前进一步。此外，本文提出的统一的判别目标函数为将基于有理函数的优化方法（比如，基于反向Jensen不等式的方法[26]）应用于MMI，MCE和MPE/MWE分类器设计中开辟了一条新路。通过采用有理函数的结构，我们期望在本文中出现的参数优化步骤里期望的算法性能都能够被确定和证明。

B. 与梯度优化的关系

GT/EBW方法和基于梯度优化的方法之间的关系在一些文献（如[2],[53],[54]）中已经有所研究 。除了收敛性方面的主要区别，基于GT/EBW的更新公式(54)的学习速度和二次牛顿修正有一定可比性；也就是说，它可以用步长的梯度上升来表示，逼近反向Hessian算子H的目标函数。在CDHMM情况下，对于目标函数的式(25)形式，我们以平均向量的估算作为例子。$O(\Lambda)$w.r.t. μ_i的梯度能表达如下：

$$\nabla_{\mu_i} O(\Lambda)|_{\Lambda=\Lambda'} = \Sigma_i'^{-1} \sum_{r=1}^{R} \sum_{t=1}^{T_r} \Delta\gamma(i,r,t)(x_t - \mu_i') \tag{57}$$

另一方面，我们可以将GT公式(54)重写成以下等式形式：

$$\begin{aligned} \mu_i &= \mu_i' + \frac{1}{\sum_{r=1}^{R}\sum_{t=1}^{T_r}\Delta\gamma(i,r,t) + D_i} \cdot \sum_{r=1}^{R}\sum_{t=1}^{T_r}\Delta\gamma(i,r,t)(x_t - \mu_i') \\ &= \mu_i' + \frac{1}{\sum_{r=1}^{R}\sum_{t=1}^{T_r}\Delta\gamma(i,r,t) + D_i} \Sigma_i' \cdot \nabla_{\mu_i} O(\Lambda)|_{\Lambda=\Lambda'} \end{aligned} \tag{58}$$

考虑到二次牛顿修正，在降 低$\Delta\gamma(i,r,t)$和μ_i的相关性后，对于μ_i，Hessian算子H_i可由以下方程近似：

$$H_i = \nabla^2_{\mu_i} O(\Lambda)|_{\Lambda=\Lambda'} \approx -\Sigma_i'^{-1} \sum_{r=1}^{R} \sum_{t=1}^{T_r} \Delta\gamma(i,r,t)$$

因此，式(54)中GT的更新公式可以进一步写成

$$\mu_i \approx \mu_i' - \underbrace{\frac{\sum_{r=1}^{R}\sum_{t=1}^{T_r}\Delta\gamma(i,r,t)}{\sum_{r=1}^{R}\sum_{t=1}^{T_r}\Delta\gamma(i,r,t) + D_i} H_i^{-1}}_{\varepsilon_i} \nabla_{\mu_i} O(\Lambda)|_{\Lambda=\Lambda'} \tag{59}$$

它逼近二次牛顿修正$\mu_i = \mu_i' - \alpha \cdot H_i^{-1} \nabla_{\mu_i} O(\Lambda)|_{\Lambda=\Lambda'}$，通常比基于梯度的简单搜索学习速度快。其它的普遍且有效的基于梯度的优化判别训练准则的方法可参见文献[15][21][31][37][38][42][51][52][54]。

参数估计的算法收敛性对于利用判别训练准则设计的分类器而言是个重要问题。寻找分类器设计中更好的判别标准和优化方法仍然是一片非常活跃且具有生命力的研究领域。我们希望在本文中基于有理函数表示的目标函数综述能够提供其它的构造公式，并能够调动研究新学习算法发展的积极性，从而提高序列模式分类器和识别器的判别能力。

VIII. 总结

在这篇文章中，我们研究了序列模式识别中关于MMI，MCE，MPE/MWE目标函数的判别学习方法。我们提出了一种能将MMI，MCE和MPE/MWE目标函数统一为普通有理函数形式(25)的方法。对现有每种判别准则构造的有理函数结构分别进行了推导和研究。过去就已经知道MMI的有理函数形式，因此本文着重为MCE和MPE/MWE目标函数存在类似的有理函数形式提供了理论依据。另外，我们给出了MMI，MCE和MPE/MWE目标函数的有理函数形式由于常量权值因子$C_{DT}(s_1 \ldots s_R)$而不同，且这些权值因子只与标记序列$s_1 \ldots s_R$有关，与被优化的参数集合Λ不相关。

为MMI，MCE和MPE/MWE推导出的有理函数形式允许基于GT/EBW的参数优化框架直接应用于判别学习中。在过去，缺乏合适的有理函数形式对于MCE和MPE/MWE是一个难题，因为如果没有这个形式，基于GT/EBW的参数优化框架就不能直接被应用。在指导性风格中，我们为判别学习中使用MMI，MCE和MPE/MWE准则而对离散HMMs和CDHMMs分别推导出基于GT/EBW的参数优化公式。

本文的整体性综述有大量被引用和研讨的早期文献贯穿全文。这里我们提供一个关于背景工作的简要总结。GT技术最早出现在文献[3]和[4]中，用于离散HMMs的ML估计，在文献[18]中得到扩展以用来解决HMMs的MMI估计问题。论文[41]将文献[18]中的研究工作从离散HMMS扩展到在小规模语音识别任务中的连续高斯分布CDHMMs。在文献[59]和[60]中扩展到大规模语音识别任务的实现。论文[47]进一步将MMI准则提高至MPE/MWE准则。同时，文献[20]中的研究工作为论文[41]的实现 提供了一个可选方案，尝试提供更严格的MMI目标函数正增长的CDHMM模型重估公式。此尝试的一个关键错误被文献[2]纠正，因而得以建立该正增长存在的依据。在判别学习发展的另一条道路上，MCE的最初形式出现在文献[1]中，而在文献[28]和[29]中得到充分发展，展示了它如何被应用于语音识别。文献[22]和[23]中的工作进一步表明，将使用文献[20]中的公式取代文献[41]中的公式，那么GT也能够应用于MCE。最近，文献[37]和[38]中的研究工作证明了MCE能应用于大规模语音识别 。再近些，文献[52]验证了非序列的梯度方法也能被成功地应用于CDHMMs中的MCE学习上。再近一些，文献[35],[53]和[54]中的研究工作表明MPE,MCE和MMI是因目标函数的普遍、非有理函数描述而关联。最后，在本文中，我们阐述了所有MMI,MCE和MPE/MWE都能严格地化为一个满足严格GT类优化的有理函数公式。

这篇文章是由近来语音识别领域基于MMI，MCE和MPE/MWE判别准则的斐然成绩所感而发。在过去，尽管MMI（诞生于1983年，见文献[39]），MCE(诞生于1967年，见文献[1])，MPE/MWE(诞生于2002年，见文献[45])具有相对较长的发展史，但缺乏对这些技术间相互关系的共同理解。由于这些技术的复杂性和共同的根本的理论主题和结构的缺失，各种判别学习发展各不相同，且参数优化也变成一个重要课题。这篇文章的主要目标是在目标函数级别上为MMI,MCE和MPE/MWE提供根本基础支撑，使新的参数优化技术得以发展，引入其它的模式识别概念，比如判别边界[66]的概念，到当前的判别学习范例中。

附录: 计算GT公式中的Δγ(i,r,t)

在(41)中，直接计算 $\Delta\gamma(i,r,t)$ 非常复杂，需要进一步简化。以下，我们设定 $s'=s_1\ldots s_{r-1}$，$s''=s_{r+1}\ldots s_R$，$X'=X_1\ldots X_{r-1}$，及 $X''=X_{r+1}\ldots X_R$.于是，由(41),可得:

$$\Delta\gamma(i,r,t)=\sum_{s_r}p(s_r|X_r,\Lambda')\times\left[\sum_{s'}\sum_{s''}p(s',\underbrace{s''|X',X'';\Lambda')(C(s',s_r,s'')}_{\Psi}-O(\Lambda'))\right]\gamma_{i,r,s_r}(t) \tag{60}$$

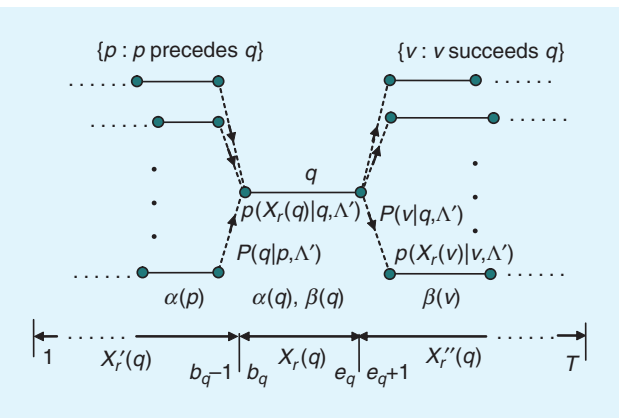
下面针对MMI和MCE/MPE/MWE讲述具体方法。

A. 对乘积形式的C(s)（如MMI）

对MMI，$C(s)=C(s_1,\ldots,s_R)=\prod_{r=1}^{R}C(s_r)=\prod_{r=1}^{R}\delta(s_r,S_r)$ 为乘积形式，于是

$$\Delta\gamma(i,r,t)=O(\Lambda')\left[\gamma_{i,r,S_r}(t)-\sum_{s_r}p(s_r|X_r,\Lambda')\gamma_{i,r,s_r}(t)\right] \tag{61}$$

当用识别字网格(lattice)时，图一演示了一个高效的算法。具体方法见文献[24]。



[FIG1] Illustrations of the sub-lattice that contains arc *q* and of the computation of the forward and backward (*q*) and (*q*) based on the sub-lattice. Each solid line represents an arc in the sub-lattice, and each dashed line represents the direct connection between two arcs (i.e., $b_q-1=e_p$).

B. 对求和形式的C(s)（MCE和MPE/MWE）

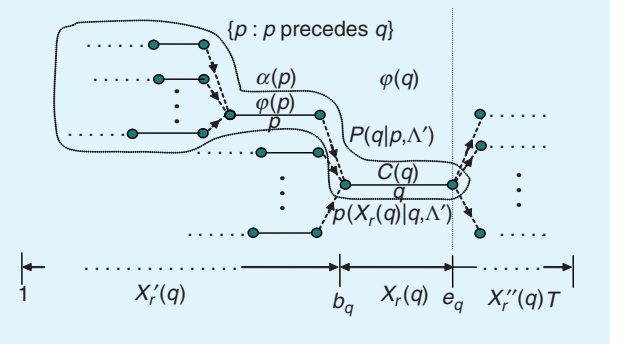
对MCE和MPE/MWE，$C(s)=C(s_1,\ldots,s_R)=\sum_{r=1}^{R}C(s_r)$，或者$C(s',s_r,s'')=C(s_r)+C(s',s'')$.于是

$$\Delta\gamma(i,r,t)=\sum_{s_r}p(s_r|X_r,\Lambda')\left(C(s_r)-\frac{\sum_{s_r}p(X_r,s_r|\Lambda')C(s_r)}{\sum_{s_r}p(X_r,s_r|\Lambda')}\right)\gamma_{i,r,s_r}(t) \tag{62}$$

对MCE，$C(s_r)=\delta(s_r,S_r)$,上述公试可简化为:

$$\Delta\gamma(i,r,t)=p(S_r|X_r,\Lambda')[\gamma_{i,r,S_r}(t)-\sum_{s_r}p(s_r|X_r,\Lambda')\gamma_{i,r,s_r}(t)] \tag{63}$$

类似的，当用识别字网格(lattice)时，图二演示了一个高效的算法。具体方法见文献[24]。



[FIG2] Illustrations of the sub-lattice containing arc *q* and of the recursive (q) computation based on the sub-lattice. Each solid line represents an arc in the sub-lattice, and each dashed line represents the transition between two arcs. The dotted line encircles all partial paths that pass *p* and end with *q*.

Agilent

Tektronix

LeCroy

Rohde & Schwarz

National Instruments

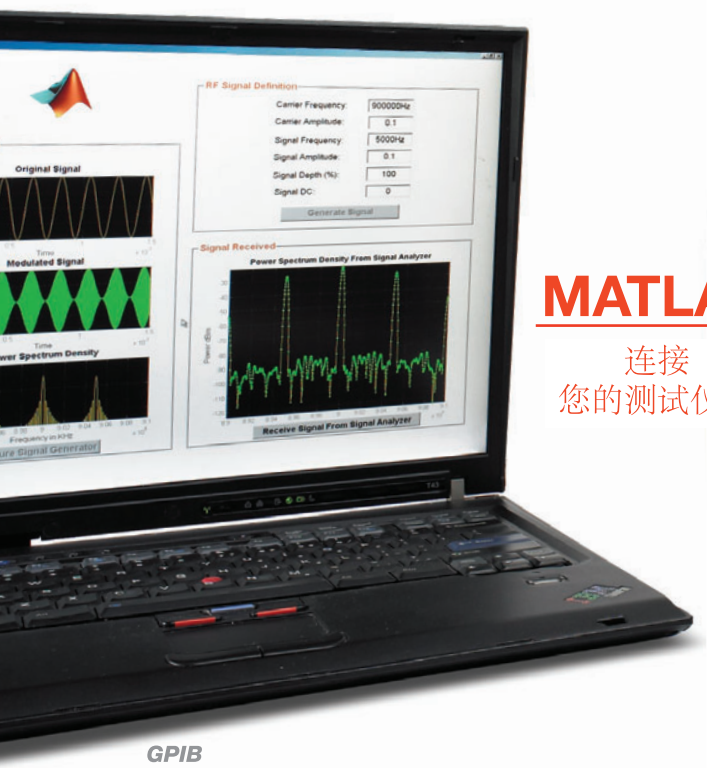
Anritsu

Keithley

Yokogawa

Tabor

Pickering



MATLAB

连接
您的测试仪器

GPIB

LXI

IVI

TCP/IP

VISA

USB

UDP

RS-232

使用标准通信协议和提供的数百种仪器驱动程序，从MATLAB®
直接连接您的测试设备。

借助于MATLAB强大的数值及绘图能力，完整分析并可视化您的测试结果。了解所支持硬件的更多信息，请访问：
www.mathworks.cn/connect

MathWorks®