信号建模与算法实践概述（I）

自我介绍

2004-2009　科大讯飞语音实验室　语音识别/增强

↓

2009-2010　科大讯飞研究院　语音识别/增强

↓

2010-2013　微软亚洲研究院　手写识别/光学字符识别

↓

2013-　中科大语音及语言信息处理国家工程实验室

团队主页：http://staff.ustc.edu.cn/~jundu/

课程信息及考核

- 主讲人：杜俊 陈雪锦 吴枫
- 助教：姜娅 康鑫
- 课程主页：http://home.ustc.edu.cn/~yajiang/
- QQ群：762742154
- 课程考核
  - 随堂小测验（20分，共4次）
  - 实验小作业（60分）
  - 团队大作业（20分）

课程知识架构

应用层	语音信号、图像信号、生物信号等等
中间层	各种模型工具：GMM 、 HMM、 NN等等
底层	微积分数理统计 、线性代数、优化方法等等

理论课时初步安排

- 信号建模与算法实践概述
  - 多模态技术应用
  - 模式分类概述：特征提取
- 数学基础知识回顾
  - 概率论，信息论，函数优化
- 模式分类
  - 贝叶斯决策理论，区分性模型，产生式模型
- 应用
  - OCR，语音增强，人脸检测/识别，图像识别

实验课时初步安排

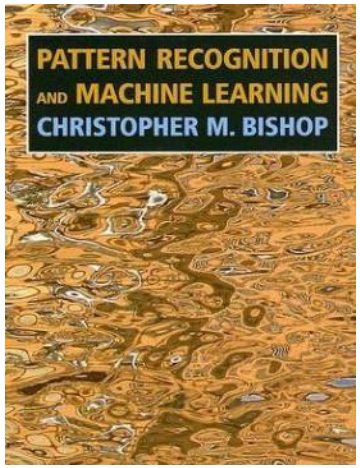
- 实验小作业
  - 实验1：语音特征提取
  - 实验2：图像处理基本操作
  - 实验3：基于单高斯模型的二分类
  - 实验4：基于多高斯模型的二分类
  - 实验5：图像分割
  - 实验6：人脸特征提取
- 团队大作业（不超过3人）
  - 建议尽早组队
  - 2选1：声学场景分类和手写数字识别
  - 期末团队答辩

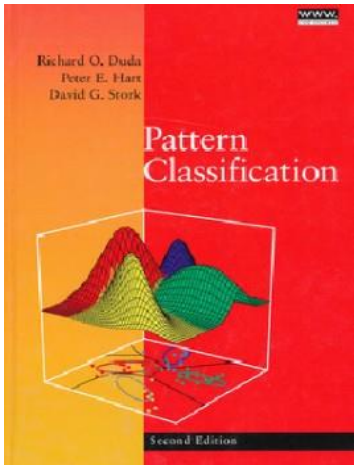
课程定位

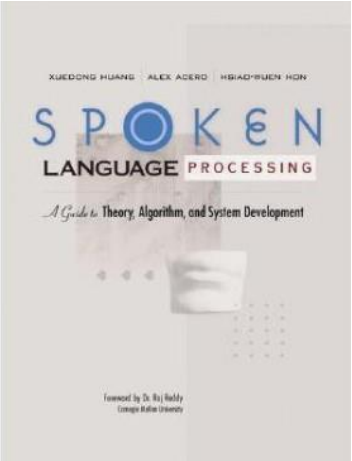
- 注重应用层面，不追求知识点面面俱到
- 注重概念介绍，不深入探讨细节
- 结合热门应用，对常用数学工具进行重新认识和学习
- 结合实例，认识到工程化能力的重要性

参考资料

- 教材相关







- 课后读物
- Wikipedia, Coursera
- 预备知识
  - 概率统计相关
  - 线性代数和矩阵分析
  - 常用编程语言：C/C++, Matlab, Python

概述

- 多模态技术应用简介
  - 语音识别
  - 手写识别
  - 光学字符识别（OCR）
  - 人脸识别
  - 图像识别

- 模式分类
  - 总体架构
  - 特征提取

自然用户界面（NUI）

多模态技术相关子方向

- 语音编码（Speech Coding）
- 语音增强（Speech Enhancement）
- 语音合成（Speech Synthesis）
- 语音识别（Speech Recognition）
- 语种识别（Language Identification）
- 说话人识别/确认（Speaker Identification/Verification）
- 手势识别（Gesture Recognition）
- 手写识别（Handwriting Recognition）
- 光学字符识别（Optical Character Recognition）
- 人脸识别（Face Recognition）
- 图像识别 (Image Recognition)

什么是语音识别

- 人与人语音交流时，识别无处不在

- 让机器听懂人说话
  - 最基本的方式：从语音到文本
  - 更高级的方式：从语音到理解

- 语音识别的重要意义
  - 最自然快捷的人机交互方式之一
  - 残疾人士：手和眼睛有使用障碍时

语音识别领域发展现状

- 产业界
  - Nuance：全球最大的语音识别技术提供商
  - IBM：具有强大数学底蕴的老牌语音识别研究机构
  - Microsoft: Windows Phone和Kinect
  - Google: GOOG-411电话查询，语音搜索
  - 科大讯飞，中科信利
  - Facebook，Amazon，BAT，华为，小米，京东，字节
- 学术界
  - Cambridge：HTK工具对学术界研究推动巨大
  - CMU：SPHINX系统
  - JHU，GIT，MIT
  - 中科大，清华，北大，声学所，自动化所
  - 上交，西工大，天大
- 关于实用化的两种论调

语音领域会议及期刊

- 国际会议
  - ICASSP
    International Conference on Acoustics, Speech and Signal Processing
  - INTERSPEECH
    Annual Conference of the International Speech Communication Association
  - ASRU
    IEEE Automatic Speech Recognition and Understanding Workshop
  - ISCSLP
    International Symposium on Chinese Spoken Language Processing

- 顶级期刊
  - IEEE Transactions on Audio. Speech and Language Processing
  - Computer Speech and Language
  - Speech Communication

语音识别传统应用

- 电脑桌面应用（听写机）
  - IBM ViaVoice，Microsoft Whisper，Nuance Dragon
- 通信应用（自动呼叫中心）
- 医疗应用（语音转录）
- 车载应用（语音导航）

- 传统应用早期在国内没有普及
  - 识别率不够，无法达到用户预期
  - 国内人工成本低，无法普及

语音识别新兴应用（移动终端）

- 语音搜索

Google™ Microsoft® Baidu 百度 www.baidu.com

- 语音助理
  - 苹果Siri
  - Google Now
  - 讯飞语点

语音识别新兴应用

- 手机输入法
  - 讯飞输入法用户已过亿

- 可穿戴式设备
  - 谷歌眼镜
  - 苹果手表

- 语音到语音的翻译

- 多模态语音识别

语音应用爆发原因

语音合成、识别等基础技术已经成熟，达到实用门槛

移动终端、 无线网络、云平台等环境条件基本完备

语音识别难点

- 人是天生就能听懂别人说话吗？
  - 每个人都要经历学习语言的过程

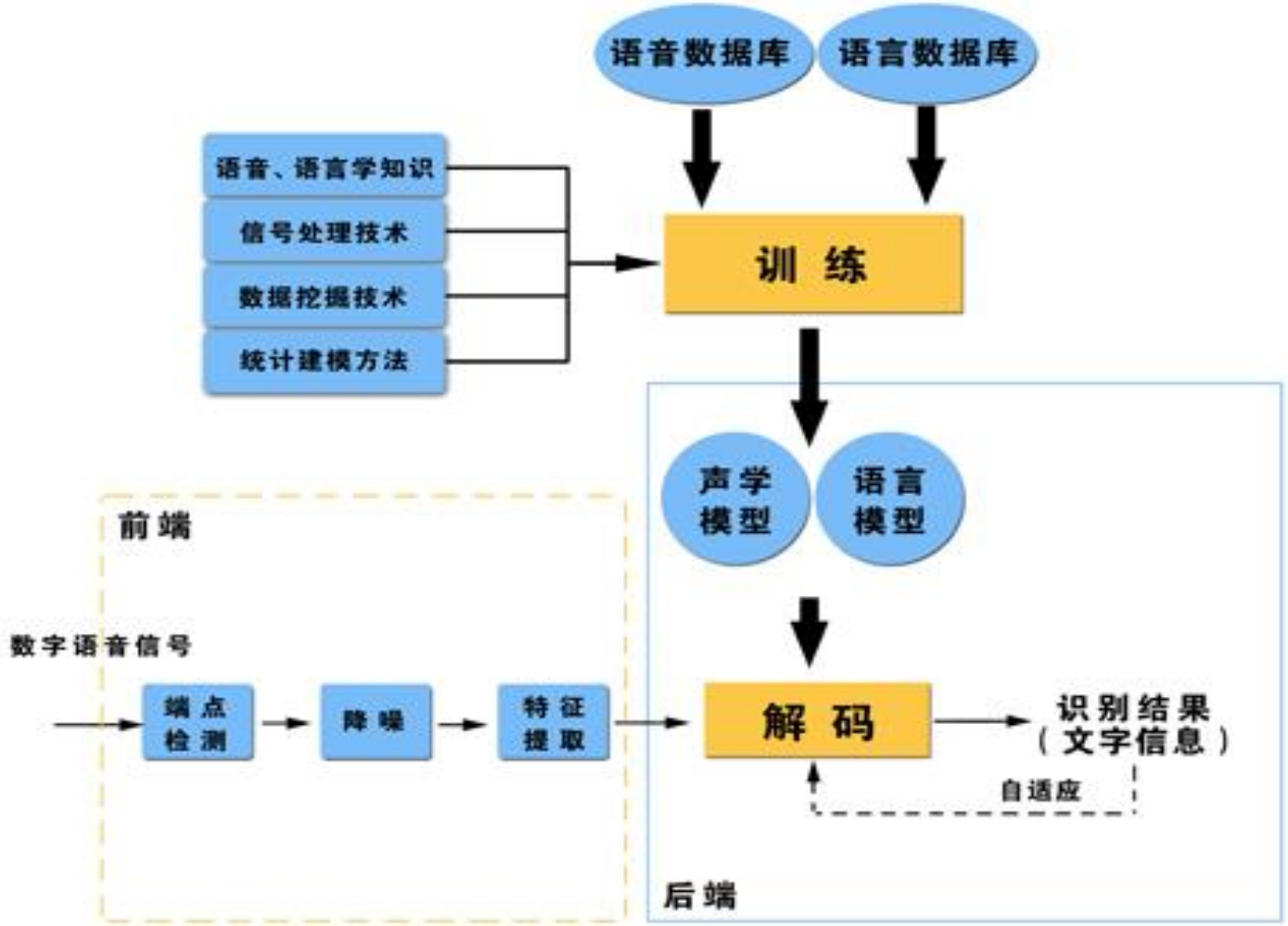
- 人在什么情况下听不懂（清）别人说话？
  - 外国语言：听不懂，需要重新学习
  - 方言口音：听不清，发音发生变化
  - 语速过快：听不清，发音不够饱满
  - 噪声环境：听不清，被噪声所掩盖

- 机器同样存在这些问题

机器和人的差距

识别任务	词汇量	机器识别错误率(%)	人识别错误率(%)
连续数字串	11	0.5	0.009
华尔街日报(干净)	5000	3	0.9
华尔街日报(带噪)	5000	9	1.1
Switchboard	65000	20	4

语音识别系统总体框架


语音数据库
语言数据库
语音、语言学知识
信号处理技术
数据挖掘技术
统计建模方法
训 练
声学模型
语言模型
前端
数字语音信号
端点检测
降噪
特征提取
解 码
识别结果（文字信息）
自适应
后端

手写识别

屏幕小 → 屏幕大

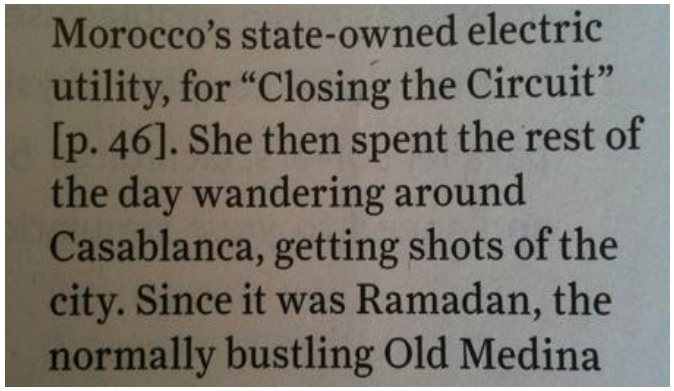
单字/叠写

单行连写

任意写

光学字符识别（OCR）

- 传统场景OCR



- 新场景OCR

文字检测和文字识别

M-431
Subway
MTA 1 2 3
CAMERA COMPUTERS
ELECTRONICS
VOL. MULTI REGION INTERNATIONAL
VENTORY CLEARENCE
25% - 50%
sbarro
ATM
MADMEN

手写识别和OCR领域发展现状

- 产业界
  - ScanSoft（Nuance），ABBYY
  - Vision Objects，Quest Visual， Vicarious
  - 汉王，微软亚洲研究院
  - 摩托罗拉（上海），科大讯飞
  - Google, BAT，华为，美团，字节

- 学术界
  - 北大，清华，自动化所，华南理工，华科，中科大

手写识别和OCR会议及期刊

- 国际会议
  - ICDAR
    International Conference on Document Analysis and Recognition
  - ICPR
    International Conference on Pattern Recognition
  - ICFHR
    International Conference on Frontiers in Handwriting Recognition
- 顶级期刊
  - PAMI
    IEEE Transactions on Pattern Analysis and Machine Intelligence
  - PR
    Pattern Recognition
  - IJDAR
    International Journal on Document Analysis and Recognition

手写识别应用

- MyScript Notes (From Vision Objects)

- 基于Kinect的空中手写系统

Leap Motion

OCR应用

- Word Lens （iPhone App from Quest Visual ）
  - 增强现实

- Translator （Windows Phone App)
  - 国外旅行必备

OCR应用

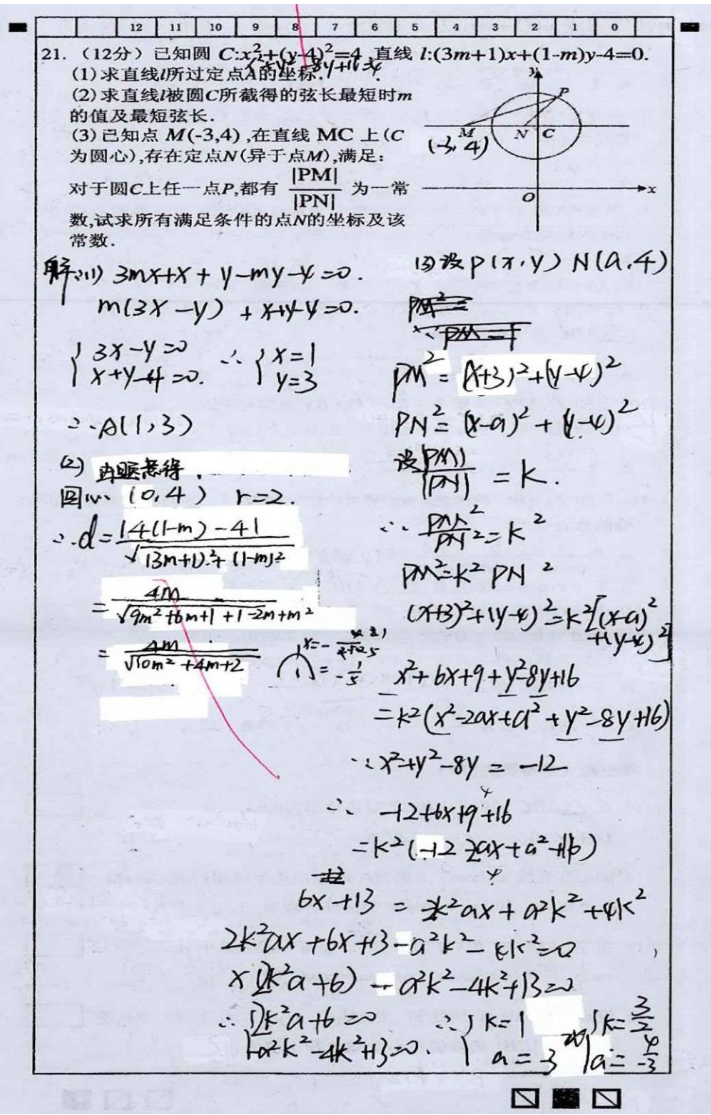
- Vicarious的反验证码技术
  - Recursive Cortical Network

手写识别和OCR难点

- 手写识别
  - 单字识别：如何进行用户自适应
  - 连续识别：如何让切分更加准确
  - 自由识别：如何矫正旋转变换

- OCR

手写识别和OCR难点



21. （12分）已知圆 $C:x^2+(y-4)^2=4$，直线 $l:(3m+1)x+(1-m)y-4=0$.

(1) 求直线 l 所过定点 A 的坐标.

(2) 求直线 l 被圆 C 所截得的弦长最短时 m 的值及最短弦长.

(3) 已知点 $M(-3,4)$，在直线 MC 上（C 为圆心），存在定点 N（异于点 M），满足：对于圆 C 上任一点 P，都有 $\frac{|PM|}{|PN|}$ 为一常数，试求所有满足条件的点 N 的坐标及该常数.

解：(1) $3mx+x+y-my-4=0$.

$m(3x-y)+x+y-4=0$.

$\begin{cases}3x-y=0\\x+y-4=0\end{cases}$ ∴ $\begin{cases}x=1\\y=3\end{cases}$

∴ $A(1,3)$

(2) 由题意得，圆心 $(0,4)$ $r=2$.

∴ $d=\frac{|4(1-m)-4|}{\sqrt{(3m+1)^2+(1-m)^2}}$

$=\frac{4m}{\sqrt{9m^2+6m+1+1-2m+m^2}}$

$=\frac{4m}{\sqrt{10m^2+4m+2}}$

$x=-\frac{b}{2a}=-\frac{1}{5}$

(3) 设 $P(x,y)$ $N(a,4)$

$PM^2=(x+3)^2+(y-4)^2$

$PN^2=(x-a)^2+(y-4)^2$

设 $\frac{|PM|}{|PN|}=k$.

∴ $\frac{PM^2}{PN^2}=k^2$

$PM^2=k^2PN^2$

$(x+3)^2+(y-4)^2=k^2[(x-a)^2+(y-4)^2]$

$x^2+6x+9+y^2-8y+16$

$=k^2(x^2-2ax+a^2+y^2-8y+16)$

∵ $x^2+y^2-8y=-12$

∴ $-12+6x+9+16$

$=k^2(-12-2ax+a^2+16)$

$6x+13=k^2ax+a^2k^2+4k^2$

$2k^2ax+6x+13-a^2k^2-4k^2=0$

$x(2k^2a+6)-a^2k^2-4k^2+13=0$

∴ $\begin{cases}2k^2a+6=0\\a^2k^2-4k^2+13=0\end{cases}$ ∴ $\begin{cases}k=1\\a=3\end{cases}$ 或 $\begin{cases}k=\frac{3}{2}\\a=-\frac{10}{3}\end{cases}$