1. 媒体与认知概述

媒体：信息的载体。信息：物质相互作用中反映出的事物属性与状态。

媒体的分类：感觉媒体，表示媒体，显示媒体，存储媒体，传输媒体

数字媒体的生命周期：获取，存储，传输，分析，处理，显示，利用

信号：信息的物理层次载体

数据：反映客观事物运动状态的信号通过感觉器官或观测仪器感知，形成了文本、数字、事实或图像等形式的数据

知识：人们在改造世界的实践中所获得的认识和经验的总和，是可用于指导实践的信息

信息是熵的减少

信息熵$H(X)=-\sum_{k=1}^{K}p(x_k)\log p(x_k)$

多元正态分布随机变量的信息熵$H(x)=(d/2)\log(2\pi e)+0.5\log|\Sigma|$

信息必须经过加工、处理才能被认知

人的认知：包括感知、注意、记忆、学习、思维、意识、情绪等。分为视觉认知、听觉认知、语言认知。

人工智能：符号主义，连接主义，行为主义

模式识别：人在观察、认识事物和现象时，常常寻找事物和现象的相同与不同之处，根据使用目的进行分类、聚类和判断，人的这种思维能力即是模式识别

媒体与知相互作用：

获取信息：认知系统利用媒体实现主动式的信息获取，包括对客观世界，人的认知规律的认识

全面，真实，准确，时效

拓展认知：提升感知、认知能力，取代人的脑力劳动

创造新媒体：VR，AR，脑机接口等

2. 机器学习基础

机器学习：根据训练样本对系统输入输出之间依赖关系进行建模，以便在测试阶段对输入做出预测。

监督学习(有真值标签)：回归，分类　非监督学习(无真值标签)：聚类

三要素：模型定义，目标函数，求解方法

模式：模式是人们根据需要及特定的环境条件，对自然事物形成的抽象分类概念。

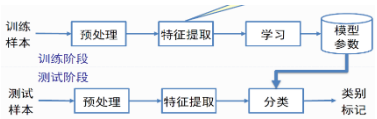
样本：样本是自然界的具体事物，具有一定的类别特性，是抽象模式的具体体现。

特征：是某一事物表现出来的特点与表征,是区别于其他事物的关键

特征提取：对输入的样本观测数据进行处理或变换，得到有利于分类的特征向量

模式识别系统：数据获取与预处理，特征提取与变换，分类决策

设计模式识别系统的阶段：训练、测试。训练使用训练集、验证集数据，测试使用测试集数据



生成式模型：对特征向量和类别的联合概率分布建模，如最小距离分类器

鉴别式模型：直接用函数对分类决策面进行建模，如神经网络

神经元模型：全或无(阈值)

人工神经元：$z=\sum_{i=0}^{n}w_ix_i \quad w_0=b,x_0=1$

Sigmoid:$f(z)=1/(1+e^{-z})$

线性回归(回归，均方误差，无激活函数)：误差反向传播和梯度下降法，随机梯度下降(SGD，速度慢，内存占用大)，批量随机梯度下降(BGD，方差大，震荡严重)，小批量随机梯度下降(MBGD，batch size=M，N 样本每 epoch，迭代次数$(n+m-1)/M$)

感知机(分类，样本点到分类决策面的距离，Sign 激活函数)

逻辑回归(分类，交叉熵，sigmoid 函数)

模型评估

容量(参数量大小)

误差(训练误差，泛化误差)：过拟合(把样本自身特性当作一般特性)，欠拟合(训练集样本的一般性质没学好)

评价指标：错误率(e/n)，识别率$(1-e/n)$

两类模式分类：

混淆矩阵：

真实类别	预测为正类	预测为负类
正类	TP	FN
负类	FP	TN

召回率$Recall=TP/(TP+FN)$，准确率$Precison=TP/(TP+FP)$

$$F_1=2\times(Precison\times Recall)/(Precison+Recall)$$

真阳性率$TPR=TP/(TP+FN)$，假阳性率$FPR=FP/(TN+FP)$

交叉验证方法：K-折交叉验证，留一法

3. 神经网络

深层神经网络(深度学习)技术的兴起得益于大数据与并行计算能力

激活函数：

Sigmoid：饱和区梯度消失，通常应用在 DNN 最后一层

Tanh：饱和区梯度消失

ReLU：正值部分梯度有效传递，解决梯度消失问题。

感知机是线性分类模型，不能求解线性不可分问题——增加一个隐含层的多层感知机

在输入层和输出层之间增加一个或多个隐含层，每层有多个神经元节点构成，多层感知机，即前馈神经网络。

具有一个隐含层的网络可模拟任一凸多边形或无界的凸区域.即隐含层每个神经元拟合凸多边形的一条边。

Softmax：$softmax(z)_i=\exp(z_i)/\sum_{j=0}^{C}\exp(z_j)$

交叉熵：$-\log(q_i)$

改进的 Softmax：$softmax(z)_i=\exp(z_i+B)/\sum_{j=0}^{C}\exp(z_j+B)$

4. 深度学习

深度学习是基于人工神经网络的机器学习技术，通过构建具有多个隐含层的 深层神经网络来实现数据驱动的模型参数学习。

基本深度学习框架：DNN，CNN，RNN

优化方法(避免陷入局部极小点)：

学习率衰减，如分段衰减，指数衰减。

对于 SGD($\theta_t=\theta_{t-1}-\eta\cdot\nabla_\theta L_t(\theta)=\theta_{t-1}-\eta g_t$)：

梯度估计调整：动量法：$\Delta\theta_t=\rho\Delta\theta_{t-1}-\eta g_t$

梯度估计调整+自适应学习率：Adam

设计有助于优化的模型：使用更多线性函数，在各层之间增加跳跃连接，增加隐藏层辅助目标函数

参数初始化：权值参数不能全部初始化为 0。预训练初始化，随机初始化，固定值初始化，高斯分布初始化，均匀分布初始化

超参数：层数，每层神经元个数，学习率，正则化系数，mini-batch 大小

数据预处理：图像大小归一化，数值动态范围归一化(最大最小值归一化，标准化，PCA)，**如果不同维度数据范围相差过大，梯度更新方向与最优解的位置相差很大。**

正则化方法(防止过拟合)：L1 或 L2 正则化，权值衰减(在标准 SGD 中，与 L2 正则化效果相同)，Dropout，提前终止，数据增强

5. 卷积神经网络原理

包含一个或多个卷积层、池化层（或称亚采样层）、全连接层

分类任务：Softmax+交叉熵

卷积层：每个神经元的输入为前层输出的局部区域，通过卷积计算提取该局部区域的特征。

池化层：对卷积层输出的特征图进行亚采样，在保留有用信息的基础上减少数据处理量，实现信息的汇聚。

局部感受野：影响元素 x 的前向计算的所有可能输入区域叫做 x 的感受野。

权值共享：同一个权值的多次使用（**全连接层没有用到权值共享**）。

卷积的动机：稀疏连接，上层的单元可以间接的连接到全部或者大部分输入图像。

卷积核尺寸：长、宽、输入通道数(Channel)。

其他常用超参数：移动步长(Stride)、边界延拓(Padding)、空洞卷积的膨胀率(Dilation)、分组数目。

输出：$A_o=(A_{in}-K_W+2P)/S+1, A=W\ or\ H$

参数量：$K_H\times K_W\times C_{in}\times C_{out}$

特殊卷积

空洞卷积：增大感受野尺寸，无空洞时 dilation 为 1。

分组卷积：在深度上进行划分，即某几个通道编为一组，相应的，卷积核深度等比例缩小而大小不变；利用每组的卷积核同他们对应的组内的输入数据卷积，得到了输出数据以后，再进行组合。

深度可分离卷积：每一个通道用一个 filter 卷积之后得到对应一个通道的输出，然后再进行信息的融合。减少了所需参数，实现了通达和区域分离。

可变形卷积

Batch Normalization：逐层对数据进行归一化，缓解梯度消失

$$y=\left[(x-\mu)/\sqrt{\sigma^2+\varepsilon}\right]*\gamma+\beta$$

其中 γ,β 是需要学习的参数，维数为通道数

训练完成时：计算整个数据集上的均值 $\boldsymbol{\mu}$和方差 $\boldsymbol{\sigma}$，用于测试过程

测试过程：每个批次计算所用的均值和方差采用训练完成时得到的统计量

常见 CNN 模型：LeNet（手写数字识别，大数据+深度卷积神经网络）AlexNet（图像分类，GPU 训练，百万量级数据） VGGNet（很深，使用 3*3 卷积核） GoogLeNet（Inception 结构，1*1 卷积用于特征降维。利用池化后的特征图进行深度上的合并，拓宽网络的广度，增加网络的尺度适应性）、Inception-Net（利用辅助损失层以避免梯度消失）、ResNet（深度残差网络。针对信息传递受阻问题，在普通浅层网络中引入短接分支，转变成残差网络实现）、ShuffleNet（利用分组卷积，又利用 channel shuffle 解决全局信息流通不畅，网络表达能力不足的问题）、DenseNet（前层网络输出信息共享）、MobileNet（首先是采用 depthwise convolution 对不同输入通道分别进行卷积，然后采用 pointwise convolution 将上面的输出再进行结合）

误差反向传播：考虑权值共享

6. 卷积神经网络的应用

迁移学习：在标记样本较为充足的源域任务上预训练模型，在标记样本较少的目标域任务上进行迁移学习。

全卷积网络：图像语义分割

目标检测：单阶段(一次出所有预测框，SSD，YOLO)，两阶段(候选框→分类修正，Fast R-CNN)

交并比：测量两个框的相似性 $J=|A\cap B|/|A\cup B|$

7. 循环神经网络

序列建模任务：语音识别，手写文字识别，语言模型，机器翻译等

其他序列建模方法：隐含马尔可夫模型

传统循环神经网络：反馈的目的是通过输出对输入的影响来改善系统的运行状况及控制效果

Elman 网络：将上一时刻隐含层输出作为隐含层的反馈输入

Jordan 网络：将上一时刻的网络输出作为隐含层的反馈输入。

BPTT 算法：

$$\frac{\partial L}{\partial U}=\sum_{t=1}^{T}\frac{\partial L}{\partial U_t}=\sum_{t=1}^{T}\sum_{s=t}^{T}\frac{\partial L_s}{\partial U_t},\qquad \frac{\partial L}{\partial x_t}=\sum_{s=t}^{T}\frac{\partial L_s}{\partial x_t}=\sum_{s=t}^{T}\frac{\partial L_s}{\partial h_t}\frac{\partial h_t}{\partial x_t}$$

截断的 BPTT：只需在有限时间内反向传播误差。

传统循环神经网络的缺点：梯度消失($\partial h_{i+1}/\partial h_i<1$),梯度爆炸($>1$)

门控机制循环神经网络：缓解梯度消失，有利于学习长距离相关信息

$$gate=\sigma(w_x^Tx_t+w_h^Th_{t-1}+b),\ \sigma=sigmoid$$

长短时记忆单元 LSTM(输入门、输出门、遗忘门、输入节点)，门控循环单元 GRU(更新门、重置门)，两者都用 sigmoid。大部分情况下是 LSTM 效果好，但 GRU 参数少。

深层循环神经网络两种常用架构 Bi-LSTM+CTC，encoder-decoder

8. 注意力机制

提高模型描述能力，解决信息超载问题。

注意力系数：$x_i,q\in\mathbb{R}^d$，计算向量相似度：

加性模型：$s(x_i,q)=\boldsymbol{u}^T\tanh(W_xx_i+W_qq)$

点积模型：$s(x_i,q)=x_i^Tq$

缩放点积模型：$s(x_i,q)=x_i^Tq/\sqrt{d}$

双线性模型：$s(x_i,q)=x_i^TWq$

深度学习中的注意力机制：在特征表示中引入加权系数

卷积神经网络中的注意力机制：

基于空间位置的注意力机制：Hard Attention(基于空间位置选取图像中一个合适的子区域再进行后续处理);Soft Attention(例如，空间位置掩膜作为注意力加权系数)

基于通道的注意力机制：Squeeze-and-Excitation Net

Squeeze：各层分别进行全局池化($W\times H\times C\to 1\times 1\times C$)

Excitation:：再通过两层全连接网络以及中间的 ReLU 激活层，最后通过 Sigmoid 函数输出维度尺寸为 $1\times 1\times 1000$ 的权重向量，分别乘以对应的 1000 个通道的特征图作为输出。

自注意力机制

序列数据时序局部依赖关系建模——如何建立非局部的依赖关系：全连接层无法处理变长问题——连接权重α_{ij}由注意力机制动态生成，动态进行"全连接"。

输入序列为 X，由 X 生成 Query，Key，Value 即 Q，K，V

计算网络中的隐含表示 H：$h_i=att((K,V),q_i)=\sum_{j=1}^{N}\alpha_{ij}v_j=\sum_{j=1}^{N}softmax\left(s(k_j,q_i)\right)v_j$

9. 支持向量机

方法	适用问题	模型类型	优化策略	学习算法
感知机	两类分类	鉴别式	极小化误分类点到超平面的距离	误差反向传播，梯度下降法
贝叶斯分类器	多类分类	生成式	极大似然估计，极大后验概率估计	概率计算公式
EM 算法	概率模型参数估计	-	(同上)	迭代学习
隐含马尔可夫模型	序列建模	生成式	(同上)	概率计算公式 EM算法
支持向量机	两类分类	鉴别式	间隔最大化	序贯最小优化算法 (SMO)
决策树	多类分类、回归	鉴别式	正则化的极大似然估计	特征选择，生成，剪枝
提升方法	两类分类	鉴别式	极小化加法模型的指数损失	迭代学习

支持向量机：是一种基于统计学习理论的机器学习方法

小样本条件下的统计学习方法；具备严格数学理论基础和直观几何解释；处理不均匀，离散，稀疏的数据有明显优势；适用于样本有限情况下的求解优化。

线性可分问题：最大间隔线性分类器

支持向量归一化后，间隔宽度为$2/\|w\|$

优化目标 $\text{minimize}\ \|w\|^2/2$

Lagrange 函数：$L(\boldsymbol{w},b,\alpha)=\|w\|^2/2-\sum_{i=1}^{n}\alpha_i(y_i(\boldsymbol{w^Tx_i}+b)-1)$

原问题：$\text{minimize}_{w,b}(\text{maximize}_\alpha L(\boldsymbol{w},b,\alpha))$

对偶问题：$\text{maximize}_\alpha(\text{minimize}_{w,b}L(\boldsymbol{w},b,\alpha))$

偏导：$\boldsymbol{w}=\sum_{i=1}^{n}\alpha_iy_i\boldsymbol{x_i},\sum_{i=1}^{n}\alpha_iy_i=0$

最终：$\boldsymbol{w}=\sum_{j\in SV}\alpha_jy_j\boldsymbol{x_j}$，任选一个支持向量，代入决策函数求出$b$

决策函数：$f(x;\alpha,b)=\boldsymbol{w^Tx}+b=\sum_{j\in SV}\alpha_jy_j\boldsymbol{x_j}\boldsymbol{x}+b$

近似线性可分问题：引入松弛因子

非线性分类：**核函数法**

如果需要将样本本映射到高维空间使之线性可分，采用核函数实现高维空间中的内积计算：$\Phi:\boldsymbol{x}\to\varphi(\boldsymbol{x}),\ \boldsymbol{x_i^Tx_j}=\varphi(\boldsymbol{x_i})^T\varphi(\boldsymbol{x_j})\to K(\boldsymbol{x_i},\boldsymbol{x_j})$

决策面：$f(x;,\alpha,b)=\boldsymbol{w}^T\varphi(\boldsymbol{x})+b=\sum_{j\in SV}\alpha_jy_j\varphi(\boldsymbol{x_j})\varphi(\boldsymbol{x})+b=\sum_{j\in SV}\alpha_jy_jK(\boldsymbol{x_j},\boldsymbol{x})+b$

根据核函数 K 可以得到基函数 φ，如：

$$K(x_i,x_j)=(1+x_i^Tx_j)^2=1+x_{i1}^2x_{i1}^2+2x_{i1}x_{i2}x_{j1}x_{j2}+x_{i2}^2x_{j2}^2+2x_{i1}x_{j1}+2x_{i2}x_{j2}$$

得：

$$\varphi(x_i)=\left[1,x_{i1}^2,\sqrt{2}x_{i1}x_{i2},x_{i2}^2,\sqrt{2}x_{i1},\sqrt{2}x_{i2}\right]^T$$

10. 统计模式识别

模式识别中的贝叶斯定理：特征向量$X=(x_1,x_2,\cdots,x_N)^T$，类模式$\omega_i,i=1,\cdots,C$

Bayes 公式：$p(\omega_i|\boldsymbol{x})=P(\omega_i)e(\boldsymbol{x}|\omega_i)/p(\boldsymbol{x})$

贝叶斯决策：已知先验概率 $P(\omega_i)$和类条件概率密度 $p(\boldsymbol{x}|\omega_i)$，计算后验概率$p(\omega_i|\boldsymbol{x})$，选择后验概率最大的类，可实现最小错误率判决 $\omega(\boldsymbol{x})=\arg_{i=1,2,\cdots,C}\max p(\omega_i|\boldsymbol{x})=\arg_{i=1,2,\cdots,C}\max p(\boldsymbol{x}|\omega_i)p(\omega_i)$

判别函数$g_i(x)=p(\omega_i)p(\boldsymbol{x}|\omega_i)$，如果$g_i(x)>g_j(x),\forall j\neq i$，将 x 归入 ω_i 类

两类问题的决策面方程：$g(x)=g_1(x)-g_2(x)$

正态分布条件下的贝叶斯决策

$$p(x|\omega_i)=\left(1/(2\pi)^{d/2}|\Sigma_i|^{1/2}\right)\exp\{-(x-\mu_i)^T\Sigma_i^{-1}(x-\mu_i)/2\}$$

最大似然估计：$l(\theta)\equiv\ln p(D|\theta)=\sum_{i=1}^{n}\ln p(x_i|\theta)$，$\hat{\theta}=\arg\max_\theta l(\theta)$

均值的最大似然估计：$\hat{\mu}=\hat{\theta}_1=\sum_{k=1}^{n}x_k/n$

协方差矩阵的最大似然估计：$\hat{\Sigma}=\hat{\theta}_2=\sum_{k=1}^{n}(x_k-\hat{\mu})(x_k-\hat{\mu})^T/n$

协方差矩阵的无偏估计：$\Sigma_{unbiased}=\sum_{k=1}^{n}(x_k-\hat{\mu})(x_k-\hat{\mu})^T/(n-1)$

判别函数 $G_i(x)=p(x|\omega_i)p(\omega_i)$，其中$p(x|\omega_i)$为上述正态分布

取对数得：$g_i(x)=-(x-\mu_i)^T\Sigma_i^{-1}(x-\mu_i)/2-d\ln(2\pi)/2-\ln(|\Sigma_i|)/2+\ln(p(\omega_i))$

在各类的协方差矩阵相同时，正态分布下的贝叶斯决策成为线性分类器。

假设各类概率相等：①$\Sigma_i=\sigma^2I$，**最小欧式距离分类器**；②$\Sigma_i=\Sigma$，**最小马氏距离分类器** $d=\sqrt{(x-\mu_i)^T\Sigma_i^{-1}(x-\mu_i)}$；③$\Sigma_i\neq\Sigma_j$，**二次判别函数**。前两者相当于线性判别函数，后一种相当于非线性判别函数。

混合高斯模型：

聚类：将样本分为 k 个子集，分别计算均值m_i，误差平方和准则：$J_e=\sum_{i=1}^{k}\sum_{x\in C_i}\|x-m_i\|^2$

K-means 聚类方法：输入聚类中心数 k 与 n 个样本，输出 k 个聚类中心，使平方误差准则最小。采用迭代的方法，确定初始中心，然后划分样本点到各自聚类，然后重新计算各聚类中心，然后重复上述步骤，直到聚类中心不再变化。必须在平均值被定义情况下使用，且初值不同，结果可能不同，对于"噪声"和孤立点敏感。

混合高斯模型，最大化 EM 算法。

11. 隐含马尔可夫模型

解决序列建模问题

马尔可夫过程是无后效性的随机过程。将来的状态仅决定于现在的状态，而和过去的状态无关，即马尔可夫性。参数集和状态空间都是离散的马尔可夫过程称为马尔科夫链。

隐含马尔可夫模型：是一个双重随机过程，状态序列是马尔科夫链，用转移概率描述，每个状态对应一个可以观测的时间，用观测概率描

述

基本要素：用三元组$\lambda=(\pi,A,B)$描述，A为转移概率矩阵，B为观测值概率分布，π为初始状态空间的概率分布

基本假设：

状态序列$q_1,\cdots,q_T$的马尔可夫性$P(q_{i+1}|q_i,\cdots,q_1)=P(q_{i+1}|q_i)$

齐次性：状态转移概率和具体时刻无关 $P(q_{i+1}|q_i)=P(q_{j+1}|q_j)$

观测序列的独立性$P(O_1,\cdots,O_T|q_1,\cdots,q_T)=\prod_{t=1}^{T}P(O_t|q_t)$

三个问题：评估(由观测序列和模型算概率$P(O|\lambda)$)，解码(根据观测序列和模型计算最优状态序列)，学习(根据样本集合的观测序列对模型的参数进行估计)

评估问题：

直接计算方法：$P(O|\lambda)=\sum_q P(O|q,\lambda)P(q|\lambda)$

引入前向、后向辅助变量$\alpha_t(i)=P(o_1,\cdots,o_t,q_t=S_i|\lambda),\beta_t(i)=P(o_{t+1},\cdots,o_T|q_t=S,\lambda)$

方法一：前向计算

初始化：$a_1(i)=\pi_i b_i(O_1)$;

递归：$\alpha_{t+1}(j)=b_j(O_{t+1})\sum_{i=1}^{N}\alpha_t(i)a_{ij},1\le t\le T-1,1\le j\le N$

算法终止时：$P(O|\lambda)=\sum_{i=1}^{N}\alpha_T(i)$

方法二：后向计算

初始化$\beta_T(i)=1,1\le i\le N$

递归：$\beta_t(i)=\sum_{j=1}^{N}a_{ij}b_j(O_{t+1})\beta_{t+1}(j),\ t=T-1,\cdots,1,\ 1\le i\le N$

算法终止时：$P(O|\lambda)=\sum_{i=1}^{N}\beta_1(i)\pi_1 b_i(O_1)$

方法三：前向-后向算法

$$P(O,q_t=S_i|\lambda)=P(O_1,\cdots O_t,q_t=S_i,O_{t+1},\cdots O_T|\lambda)$$
$$=P(O_1,\cdots O_t,q_t=S_i|\lambda)P(O_{t+1},\cdots O_T|q_t=S_i,\lambda)=\alpha_t(i)\beta_t(i)$$

得到：$P(O|\lambda)=\sum_{i=1}^{N}\alpha_t(i)\beta_t(i)$

解码问题：

Viterbi 算法——动态规划

使每一时刻状态序列出现相应观测值的可能达到最大。与前向算法的区别在于前向算法累计所有路径的概率，而 Viterbi 只计算最优路径的概率。

定义$\delta_t(i)=\max P(q_1,q_2\cdots,q_t=i,o_1,o_2,\cdots o_t|\lambda)$

初始化$\delta_1(i)=\pi_i b_i(O_1),\varphi_1(i)=0,1\le i\le N$

递归
$$\delta_t(j)=\max\left[\delta_{t-1}(i)a_{ij}\right]b_i(O_t),$$
$$\varphi_t(i)=\arg_i\max\left[\delta_{t-1}(i)\,a_{ij}\right],2\le t\le T,1\le j\le N$$

终止：$P^*=\max[\delta_T(i)],q_T^*=\arg_i\max[\delta_T(i)]$

最可能的序列状态 $q_t^*=\varphi_{t+1}(q_{t+1}^*),t=1,2,\cdots,T-1$

学习问题：

求$\lambda=(A,B,\pi)$，使$P(O|\lambda)$最大

Baum-Welch 算法：是一种 EM 算法，即从不完全数据(样本特征序列与隐含状态序列的对齐关系未知)中求解模型参数的最大似然估计方法：

选择 HMM 初始参数λ_0;

求期望，即利用给定的 HMM 参数求样本特征序列的状态对齐结果；最大化：根据上一步的状态对齐结果，利用最大似然估计更新 HMM 参数 λ;

重复上述步骤，直到收敛：$\log P(O|\lambda)-\log P(O|\lambda_0)<d$

定义：$\xi_t(i,j)=P(s_t=i,s_{t+1}=j|O,\lambda)$
$=\alpha_t(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)/\sum_{i=1}^{N}\sum_{j=1}^{N}numerator$

语音识别：语音识别是机器通过识别和理解过程把人类的语音信号转变为相应的文本或命令的技术。

困难：连续语音流中各语音单位之间不存在明显的界线、语音特征变化差异大（音素的声学特征变化与上下文相关、随发音人及其生理或心理状态的变化而有很大的差异、环境噪声和传输设备的差异也将直接影响语音特征的提取）

语音的分类：按词汇量大小分、按发音方式分（孤立词识别、连接词识别、连续语音识别、关键词检出）、按说话人分（特定说话人、非特定说话人）、按识别方法分（模板匹配—利用动态时间规整 DTW 将测试语音与参考模板进行匹配、随机模型—利用 HMM 对似然函数进行估计、深度学习—利用 RNN 进行端到端学习）

奈奎斯特采样定理：$f_s>2f_{max}$

窄带语音信号：$f_s=8000Hz$，如电话语音，可以保持基本语义，不影响人对语音的感知

宽带语音信号：$f_s>16000Hz$，对语音质量要求较高的场合。

量化精度：量化所用的比特数越大，声音质量越好。人类听觉系统对声音信号强弱刺激反应不是线性的，而是成对数比例关系。

语音信号的短时分析：语音信号是一种典型的非平稳信号。语音识别中常用的帧长为 20~30ms，帧移为 10ms。短时分析做加窗处理（如汉明窗）：$S_w=Es(m)w(n-m)$。

语音信号时域分析：直接分析语音信号时域波形提取特征参数（短时平均能量、短时平均幅度、短时平均过零率、短时自相关函数）

语音信号频域分析：带通滤波器组法、傅里叶变换法、线性预测法

短时傅里叶分析（Gabor）：得到短时功率谱与语谱图。

线性预测分析：语音信号当前数据可以用过去若干数据的线性组合来预测；利用实际采样值与预测值之间的均方误差对线性组合系数进行优化求解，可得线性预测分析系数。

Mel 频率倒谱系数 MFCC：梅尔频率与实际声音频率成对数关系，对低频比高频敏感。$Mel(f)=2595\lg(1+f/700)$。

12. 特征提取与降维

特征提取：对信号空间的原始数据通过处理及变换得到在特征空间中最能反映分类本质的特征。特征提取是模式识别的重要环节，其目的是**提取类内差别小，类间差别大的鉴别力强的特征向量**（图像特征提取—Gabor 滤波器组特征等；语音特征提取—梅尔频率倒谱系数等）一维信号的时频分析：傅里叶变换、短时傅里叶变换（Gabor 变换）、小波变换。

特征降维：在给定精度下，准确地对某些变量的函数进行估计，所需样本量会随着特征维数的增加而呈指数形式增长。通过特征变换等手段，实现特征降维。克服维数灾难；获取本质特征；节省存储空间；去除无用噪声；大数据可视化。特征降维常见方法有特征选择和特征变换。

特征变换：在使优化判据J取最大的目标下，对n个原始特征进行高维空间向低维空间的映射。

主成分分析(PCA)：最小均方误差准则下保留原始数据信息。优点：采用样本协方差矩阵的特征向量作为变换的基向量，与样本的统计特性完全匹配；在最小均方误差准则下，是最佳变换。缺点：变换矩阵随样本数据而异，**无快速算法**。

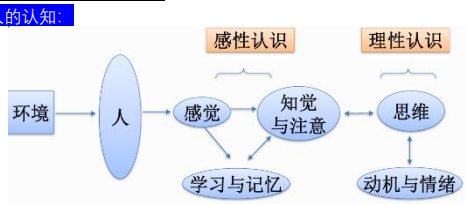
线性判别分析(LDA)：最小均方误差准则下区分多类数据。

PCA 与 LDA 的异同：相同点：降维方法，特征值分解，高斯假设；不同点：PCA 无监督/LDA 有监督；LDA 可用于分类、LDA 取分类性能最好的投影方向，PCA 取投影点方差最大的方向。

自动编码器：一种尽可能复现输入信号的神经网络；找到代表输入数据的最重要的因素。线性自动编码器等同于 PCA。

13. 认知的生物机制

人的认知：



对人认知的研究层次：分子层次（脑中最基本的成份—各种分子的功能，如神经递质、蛋白质等）细胞层次（神经元是如何行使其功能）系统层次（大量神经元构成了复杂的环路完成某一功能，如视觉感知系统、运动系统）认知层次（神经系统是如何产生认知和协调的行为）

研究认知的技术手段：脑电图（是通过电极记录下来的脑细胞群的自发性、节律性电活动）、脑成像技术（正电子发射层描记术—使用对比剂；事件相关电位&脑磁图—高时间分辨率&低空间分辨率；磁共振成像—非侵入性、没有辐射，应用广泛）、反应时间

人的行为反应：

刺激：围绕机体的一切外界因素，都可以看成是环境刺激因素，同时也可把刺激理解为信息。

行为：有机体对于所处情境的反应形式，心理学家将行为分解为刺激、生物体、反应三项因素研究。

心理实验：在严密控制的条件下，有组织地逐次变化条件，对相伴随的心理现象的变化进行观察，记录和测定，从而确定条件与心理现象之间关系的方法。

反应时间：从刺激的呈现到反应的开始之间的时间间隔（包含感觉器官、大脑加工、神经传入传出所需的时间以及肌肉效应器反应所需的时间，其中大脑加工所消耗的时间最多，STROOP 效应）。

感觉：人脑对事物的个别属性的认知。感觉提供了内外环境的信息，是人全部心理现象的基础。

知觉：将感觉信息组成有意义的对象，在已贮存的知觉知识经验的参与下，理解当前刺激的意义。对这种刺激意义的理解（获得）就是当前刺激和已贮存的知识经验相互作用的结果。

视觉感知：光刺激作用于人眼所产生。生理机制包括折光机制、感觉机制、传导机制、中枢机制。

视觉的生理机制：视网膜是眼球的光敏感层，外层有锥体细胞（主要感受物体的细节和颜色）和棒体细胞（主要感受物体的明暗）、中间有双极细胞、内层有神经节细胞。电信号从感受器产生以后，沿着视神经传至大脑。

视觉的传导机制由三级神经元实现：视网膜双极细胞（具有侧抑制作用）视神经节细胞（发出的神经纤维，经视交叉，传至丘脑外侧膝状体）第三级神经元（纤维从丘脑外侧膝状体发出，终止于大脑枕叶的纹状区）。

视觉系统的侧抑制：视网膜的双极细胞上，在神经节细胞的感受野里，在外侧膝状体以及视皮层细胞中都能产生侧抑制。侧抑制有利于视觉从背景中分出对象，尤其在看物体的边角和轮廓时会提高视敏度，使对比的差异增强。

视觉系统的中枢机制：人类的视觉皮层包括初级视皮层（V1，也称作纹状皮层）和纹外皮层（V2、V3、V4、V5 等）

感受野：视网膜上的一定区域受到刺激后会激活视觉系统中与这片区域有联系的各层神经细胞的活动。视网膜上的这个区域就是神经细胞的感受野。V1 视神经细胞主要有 3 种感受野：同心圆感受野（又称中心-周边感受野，分为 on-center 和 off-center）、简单感受野（对方向、位置和空间频率有明显选择性，为简单细胞，比较适合于检测具有明暗对比的边缘，且对边缘的位置和方位有严格的选择性。响应与 Gabor 滤波器相似）、复杂感受野（对于视觉刺激在视野中的位置没有选择性，对应于特定方向的条形刺激，具有位置不变性）。排成一条线的同心圆感受野聚合成一个简单感受野，从而对一定朝向的条形物敏感；若干个同一朝向的简单感受野，聚合到一个复杂感受野，从而使复杂感受野对任一地点的同一朝向的条形物敏感。纹外皮层具有更高级的视觉感知功能：在纹外皮层的第一个皮层区域，包含一些粗细条纹。对颜色有选择性的细胞集中在细条纹中、对运动方向有选择性的细胞则存在于粗条纹中、对形状敏感的细胞则在粗条纹和中间条纹中都有所分布。

知觉的信息加工过程：自下而上加工（数据驱动加工。是由外部刺激开始的加工，通常是先通过对较小的知觉单元进行分析，然后再转向较大的知觉单元，经过一系列连续阶段的加工，而达到对感觉刺激的解释）、自上而下加工（概念驱动加工。是由有关知觉对象的一般知识开始的加工。由此可以形成期望或对知觉对象的假设。这种期望或假设制约着加工的所有阶段或水平，从而调整特征觉察器，引导对细节的注意等）

人已有的知识和知识结构，对当前的认知活动，具有决定性作用。

注意：是人的心理活动对一定对象的指向和集中。功能包括信号检测、选择性注意、分配性注意。特征包括：选择性、持续性、注意的转移。

注意的选择性：包括指向性（在每一瞬间，其心理活动或意识选择了某个对象，而忽略了另一些对象）和集中性（当心理活动或意识指向某个对象的时候，它们会在这个对象上集中起来）

注意选择的理论模型：过滤器模型（来自外界的信息是大量的，而人的神经系统高级中枢的加工能力是有限的，于是出现瓶颈。为避免系统超载，需要某种过滤器进行调节，选择其中较少的信息，使其进入高级分析阶段，这类信息受到进一步加工而被识别和存储，其他信息则不让通过。在感觉和察觉之间进行过滤。双耳分听实验。在嘈杂的环境中，也能听到别人喊自己的名字。）、衰减模型（又称中期选择模型：高级分析水平的容量有限，必须由过滤器加以调节，不过这种过滤器不只允许关注的通道的信息通过，也允许非关注的信息衰减通过，其中一些信息仍然可得到高级加工）、反应选择模型（几个输入通道的信息均可进入高级分析水平，得到全部的知觉加工。注意不在于选择知觉刺激，而在于选择对刺激的反应，即输出是按其重要性安排的，这种安排依赖于长期的倾向、上下文和指导语。双耳同时分听的追随靶子词实验）

过滤器模型与衰减模型的比较：不同—过滤器模型假设选择性注意的基础是对进来刺激物理属性的较粗略的分析；而衰减模型则认为，前注意分析更为复杂，甚至可能由语义加工组成。过滤器理论中的过滤器是"全或无"性质，什么都未选择的通道是完全关闭的；而衰减模型则认为未选择的通道不是完全关闭的，而只是关小或阻抑。相同—两模型的根本出发点相同：高级分析水平的容量有限或通道容量有限，必须过滤器予以调节。过滤器的位置在两模型中是相同的，都处于初级分析和高级的意义分析之间。过滤器的作业又都是选择一部分信息进入高级的知觉分析水平，使之得到识别。并且，注意选择具有知觉性质，因此，二者并称为知觉选择模型。

注意的认知资源分配：双加工理论：控制性加工（受到人的意识控制与认知资源的限制，需要注意的加工。其容量有限，可灵活用于变化着的环境。习得后加工过程较难改变）自动加工（不受人所控制的加工，也不受认知资源的限制，无需应用注意。没有一定的容量限制，而且一旦形成就很难改变）经过大量的练习后，可能转变为自动加工。

注意的生理机制：朝向反射（情景的新异性引起一种复杂而又特殊的反射。它是注意的最初生理机制）、脑干网状结构（脑干网状结构的激活作用使脑处于觉醒状态，是和边缘系统和大脑皮层相联系的）、大脑皮层（大脑皮层是产生注意的最高部位）

计算机视觉中的注意力机制：注意力机制模仿了生物观察行为的内部过程，即将内部经验和外部感觉对齐从而增加部分区域的观察精细度的机制。当前时刻输出 $ht=\Sigma\alpha(xt,xt')f(xt,xt')Tt'=1$（注意力*特征）

记忆：是在头脑中积累和保存个体经验的心理过程。在信息加工的术语中，记忆是人脑对外界输入的信息进行编码、存储和提取的过程。**记忆是一个系统，具有自身结构，由三个子系统构成：感觉记忆、短时记忆、长时记忆。**

记忆的生理基础：皮层运动区—程序性记忆；额叶—语义与情节记忆；前额叶—短时记忆；颞叶—颞叶参与长时语义和情节记忆的整合与存储，对短时记忆中的新材料加工也起作用；杏仁核—新情绪记忆信息的整合；海马—整合新的长时语义和情节记忆；小脑—程序性记忆。

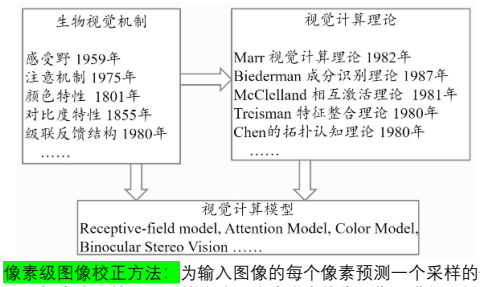
感觉记忆：又可称为瞬时记忆，是一种信息存储时间以毫秒或秒计的记忆，容量≥9。心理学家假设每一种感觉通道都有一种感觉记忆，每一种感觉记忆都能将感觉刺激的物理特征的精确表征保持几秒钟或更短的时间。感觉记忆是记忆系统的开始阶段，它是一种原始的感觉形式，是记忆系统在对外界信息进行进一步加工之前的暂时登记。特点：存储时间非常短（图像记忆在几百毫秒钟内，声像记忆可达 4 秒；信息加工只是初步的，但也可以进行信息整合；基本是按照刺激的物理特点进行编码，是外界刺激的真实复本）、记忆容量非常大（图像记忆在 9-20 个项目内，声像记忆容量小于图像记忆；但只有一部分信息会进入到高一级的短时记忆中）、记忆过程是无意识的自动化的，人无法控制。

短时记忆：一种信息存储时间为 1 分钟以内（约 15-30 秒）的记忆，又可以被称为电话号码式记忆，容量为 7±2。是个体对刺激信息进行加工、编码、短暂保持和容量有限的记忆。在短时记忆阶段，人脑同时能容纳 5-9 组内容。从短时记忆向长时记忆存入一项需要 5～10 秒钟（西蒙认为可能是 8 秒钟）。短时记忆信息的编码：在记忆系统中对信息进行转换，使之获得适合于记忆系统的形式，经过编码所产生的具体的信息形式为代码。影响因素及遗忘原因：干扰作业难度大小，记忆材料的熟悉程度；痕迹衰退说：记忆痕迹将随时间而消退；干扰说：已有信息的干扰。

长时记忆：指信息保持存储时间在一分钟以上的记忆，可以是数年甚至终生难忘，容量巨大。量的变化：存储信息的数量随时间的推移而逐渐下降。质的变化：受知识和经验差异的影响，人们存储的经验可能会发生不同程度的变化：会发生记忆的扭曲、记忆的错觉。

14. 媒体与认知相互作用

类脑认知计算



像素级图像校正方法：为输入图像的每个像素预测一个采样的偏移量和幅度值、基于预测的偏移量从该像素的临近位置进行重新采样（基于双线性插值）并乘以幅度值，从而得到校正图像。校正网络和识别网络端到端联合训练。

信息的获取与利用：被动式获取（视觉、光场相机）、主动式获取（超声探测、结构光成像）。深度信息获取：双目测距（利用双目摄像机拍摄物体，再通过视差估计以及成像几何关系计算物体距离）、TOF 飞行时间（通过专有传感器，采集近红外光从发射到接收的飞行时间，计算物体距离）结构光（结构光投射特定的光信息到物体表面后，由摄像机采集反射信号。根据物体表面变化造成的光信号的变化来计算物体的位置和深度等信息）

创造新媒体：视觉认知与媒体技术：视觉暂留与电影（视觉暂留现象视神经对物体的印象会保持 0.1-0.4 秒的时间）、立体视觉与立体显示（人脑将二维图像转化为三维图像：现代心理学认为这个复杂的处理过程分为生理学和心理学两个层面，具体又分为 10 种深度暗示，人们就是通过这些深度暗示来感知三维物体。线性透视、像的大小、重叠光照及阴影、结构梯度、面积透视；调节睫状体平滑肌、汇聚两眼的视轴、双目视差、移动视差。目前三维显示是基于双目视差原理，如果能同时提供人眼在生理学上的 4 种深度暗示，那么该技术可以称为"真"三维显示技术。）、虚拟现实与增强现实（VR：一种能够创建和体验虚拟环境的，由计算机生成的，提供多种感官刺激的自然人机交互系统。AR：在虚拟现实的基础上发展起来的新技术，也被称之为混合现实）

似动现象：两个相距不远、相继出现的视觉刺激物，呈现的时间间隔如果在 1/10 秒到 1/30 秒之间，那么我们看到的不是两个物体，而是一个物体在移动。