

默认以 e 为底。

特别说明：蓝色是幂指数或者转置操作，蓝色 \hat{a} 为 \hat{a} ，橙色为下标， π 是 π ， u 是 μ ， Σ 是协方差矩阵， σ 是标准差， $\sum\{\}$ 是求和， $|\cdot|$ 是绝对值或行列式， \cdot 是按元素乘法。如果篇幅不足，下划线部分可以删去。 \int 是积分。蓝色 $*$ 是共轭

媒认知识

第一节 概论

一、定义

1. 媒体：信息的载体
2. 信息：物质相互作用中反映出的事物属性与状态。
3. 信号：信息的物理层次载体
4. 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。
5. 知识：人们在改造世界的实践中所获得的的认识和经验的总和。

二、信息的度量

1. 信息熵： $\sum\{-p\log(p)\}$

(1) 正态分布信息熵： $0.5\log(2\pi e \sigma^2)$

(2) 多元正态分布： $p(X)=\exp\{-0.5(x-u)^T \Sigma^{-1}(x-u)\}/((2\pi)^{0.5d}|\Sigma|^{0.5})$ 。

(3) 多元正态分布信息熵： $0.5d\log(2\pi e)+0.5\log(|\Sigma|)$

2. 麦克卢汉的媒介理论：媒介即是讯息；媒介是人的延伸，可以引起人的感官能力变化，还能促进社会结构的变化。热媒介讯息清晰度高，不需要使用者过多思考，参与度低。冷媒介讯息清晰度较低，需要使用者通过自己的思考来补全缺失的讯息。

3. 媒介素养：对各种媒介信息的解读和批判能力，使用媒介信息为个人生活、社会发展所用的能力。

三、人工智能

1. 符号主义方法：认知基元是符号，只能写为通过符号操作来实现，以归结原理为基础，以 Lisp 和 Prolog 语言为代表。着重问题求解中的启发式搜索和推理过程，在逻辑思维的模拟方面取得成功，如自动证明定理和专家系统。

2. 连接主义方法：思维的基元是神经元，把智能理解为相互连接的神经元竞争与协作的结果，以反向传播网络模型和 Hopfield 网络模型更为突出。着重结构模拟，研究神经元特征、神经网络拓扑、学习规则、网络的非线性动力学性质和自适应的协同行为。

3. 行为主义方法：反馈是控制论中的基石，没有反馈就没有智能。强调智能系统与环境的交互，从运行的环境中获取信息（感知），通过行为对环境施加影响。智能行为体现在系统与环境的交互之中，功能、结构和智能行为不可分割。

第二节 机器学习基础

一、定义

1. 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。

2. 基本任务：

(1) 监督学习：训练样本有真值标签。

①回归

②分类

(2) 非监督学习：训练样本无标签。

①聚类

3. 机器学习的三要素：模型定义、目标函数、求解方法。

10 页

4. 回归问题：输入标量或向量 x ，输出随 x 变化的连续变量 y 。

5. 分类问题：输入特征向量 x ，输出类别标签 y 。

6. 模式识别基本概念

(1) 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。

(2) 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。

(3) 模式分类/模式识别：寻求样本观测量与类别属性的联系。

(4) 特征：某一事物表现出来的特点和表征，是区别于其他事物的关键。

(5) 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:14 页。

7. 生成式模型：对特征向量和类别的联合概率分布建模。

典型方法：最小距离分类器：计算样本到各类别中心的距离，选择距离最小的类别作为输出。

8. 鉴别式模型：直接用函数对分类决策面进行建模。

典型方法：神经网络。

9. 激活函数：阈值函数、符号函数、sigmoid 函数、tanh 函数。

Sigmoid 的导数 $y' = y*(1-y)$

10. 线性回归、感知机、逻辑回归

	<u>线性回归</u>	<u>感知机</u>	<u>逻辑回归</u>
<u>机器学习任务</u>	<u>回归</u>	<u>分类</u>	<u>分类</u>
<u>模型</u>	<u>$f(x)=w^T x+b$</u> <u>$w: 1 \times n$</u> <u>$b: 1 \times 1$</u>	<u>$f(x)=\text{sign}(w^T x+b)$</u> <u>$w: k \times n$</u> <u>$b: k \times 1$</u>	<u>$f(x)=\text{sigmoid}(w^T x+b)$</u> <u>$w: k \times n$</u> <u>$b: k \times 1$</u>
<u>目标函数</u>	<u>均方误差</u>	<u>样本到分类界面的距离</u>	<u>交叉熵</u>

二、线性回归方法

1. 输入 n 为向量 x ，模型有 n 个权重系数和 1 个偏置量，输出一个标量

29 页

2. 线性回归方法是一个单层单节点的网络

3. (1) 单变量线性回归：输入为标量

(2) 多变量线性回归：输入为向量

4. 模型求解：33 页

最小二乘法：找到一条直线，使所有样本点到直线上预测点的均方误差最小。

误差反向传播：梯度下降法。

梯度：用于寻找更新权重的方向。

学习率：一个在训练模型前设定的超参数，用于指定更新参数的强度。

5. 训练方法：梯度下降法调整参数，学习获取使损失函数最小化的参数。

6. 选取样本批量大小：

(1) 批量梯度下降法(BGD)：利用所有 N 个训练样本计算平均梯度。速度慢，数据量大时内存不足。

(2) 随机梯度下降法(SGD)：随机选取单个样本计算梯度。方差大，损失函数震荡严重。

(3) 小批量随机梯度下降法：每次迭代选取小批量样本计算梯度， $\text{batch size} = M$ 。 N 个样本训练一轮(epoch)的迭代次数为 $(N+M-1)/M$ 。

三、逻辑回归与分类

1. 感知机：输入样本的特征向量，输出样本类别代码，记为 1, -1。感知机对应于样本空间中的分类超平面，属于鉴别式模型。分类决策面对应于线性方程 $w^T x + b = 0$ 。

2. 感知机学习策略：损失函数定义为所有误分类点到分类决策面的等效距离之和。

误分类点：当 $y=1$ 时， $\hat{y} = w^T x + b < 0$ ；当 $y=-1$ 时， $\hat{y} = w^T x + b > 0$ ；

误分类点到分类决策面的等效距离： $-y(w^T x + b)$ (等效距离是距离的 $|w|$ 倍)

3. 感知机学习算法：梯度下降法

4. 逻辑回归的目标函数：45 页。

交叉熵通常用于度量概率分布的差异。

将类别真值看做概率分布，模型输出分类结果及对应点概率估计。

四、模型评估

1. 模型的容量：

模型复杂度：模型参数量大小等。

2. 误差：模型的实际输出与样本的对应真值之间的差异。

训练误差（经验误差）与测试误差（泛化误差）

3. 模型的选择：

拟合能力强的模型一般复杂度会比较高，容易过拟合；如果限制模型复杂度，降低拟合能力，可能会欠拟合。

(1) 过拟合：模型在训练集上过度学习，把训练样本部分自身性质当作所有潜在样本的一般形式，缺乏对问题本质的理解。表现为训练误差小，但测试误差大，出现泛化性能下降。

(2) 欠拟合：训练集样本的一般性质没学好，表现为训练误差大。

均方误差中的过拟合与欠拟合：51-52 页

过拟合对应于方差项，欠拟合对应于偏差项。

4. 评价指标：错误率与识别率

错误率：分类错误的样本占样本总数的比例。

识别率：分类正确的样本占样本总数的比例。

系统评价指标：

真实类别	预测为正类的数目	预测为负类的数目
正类	TP (真正类)	FN (假负类)
负类	FP (假正类)	TN (真负类)

$$\text{召回率: Recall} = \frac{TP}{TP+FN}$$

$$\text{准确率: Precision} = \frac{TP}{TP+FP}$$

$$\text{F1 分数: F1} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

真阳性率: $TPR = \frac{TP}{TP+FN}$

假阳性率: $FPR = \frac{FP}{TN+FP}$

5. 交叉验证方法

(1) 定义: 将原始数据进行分组, 一部分作为训练集, 一部分作为验证集。

(2) K-折交叉验证: 将原始数据分为K组(一般是均分), 将每个子集数据分别做一次验证集, 其余的K-1组子集数据作为训练集, 得到K个模型。K个模型最终的验证集的分类准确率的平均数作为分类器的性能指标。

(3) 留一法: 每个样本单独作为验证集, 其余的N-1个样本作为训练集。

第三节 神经网络

一、神经网络基本原理

1. 激活函数: sigmoid, tanh, ReLU, PReLU(Leaky ReLU) 8 页

2. 基于正态分布的线性判别函数: 线性分类决策面方程: $(u_1 - u_2)^T \Sigma^{-1} x - 0.5u_1^T \Sigma^{-1} u_1 + 0.5u_2^T \Sigma^{-1} u_2 = 0$ 大于0是第一类。

3. 多层感知机: 在输入层与输出层之间增加一个或多个隐藏层, 每层有多个神经元节点, 构成多层感知机, 即前馈神经网络。每一层的神经元节点与前一层的各结点均有线性加权连接, 称为全连接层或线性层或密集层。

4. 架构设计: 架构是指网络的整体结构。前馈神经网络包含若干层, 每一层有多个计算节点, 每一层的输出作为下一层的输入。

模型配置参数包括网络层数每一层节点数。此外, 还要指定每一层所用的激活函数, 以及是否使用偏置量。

5. 前馈神经网络的拟合能力:

(1) 非线性分类任务中拟合决策面: 二分类任务中, 假设二维特征空间中正类样本位于一个凸多边形内部, 具有一个隐藏层的网络可模拟任意凸多边形或无界的凸区域, 即隐藏层每个神经元拟合凸多边形的一条边。多层网络可模拟更为复杂的图形。

(2) 拟合逻辑运算: 具有一个隐藏层、隐藏层节点数可任意设置、激活函数采用阈值函数的网络, 可以实现任意的二值逻辑网络。

(3) 非线性回归中拟合非线性函数: 具有一个隐藏层、隐藏层节点数可任意设置、激活函数采用S型非线性函数的网络, 可以一致逼近紧集上的连续函数或按范数逼近紧集上的平方可积函数。

6. 目标函数

(1) MSE 22 页

(2) Softmax 回归

用于多类分类任务时, 输出层节点数设为模式分类问题的类别总数C, 即输出层每一节点预测属于对应类别的概率。类别真值使用 one-hot 编码表示。输出层节点激活函数采用 Softmax 函数。(23 页)

softmax 的改进: 29 页

softmax 的导数: 30 页

(3) 交叉熵: 24 页

二、神经网络的训练方法

训练时利用误差反向传播算法, 采用梯度下降法调整模型参数。

1. 误差反向传播算法：见第一次习题课
2. 正则化：防止模型在训练集上过拟合。例如，在目标函数中引入对模型复杂度的惩罚项；当验证集上误差不再降低时就提前终止训练。
3. 神经网络的非线性表示能力由一层网络中的节点数、非线性激活函数、网络层数共同决定。

第四节 深度学习

一、定义

1. 定义

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藏层的深层神经网络，来实现数据驱动模型参数学习。深度学习通过组合此等特征形成更抽象的高层特征表示，模拟大脑的任职机制来处理图像、声音、文本等数据。

2. 深度学习与传统学习方法的区别与联系

	传统方法	深度学习
特征提取	人工设计特征提取方法	自动学习特征提取方法
模型参数	样本统计分布参数等	神经网络节点的权值系数
模式分类训练准则	最小错误率等	交叉熵极小化
训练过程	多步骤、分模块或分阶段训练	端到端网络迭代训练，直接优化任务总体目标
计算规模	适中	处理数据量大，训练迭代次数多

联系：单个神经元节点与传统线性分类器模型类似，深度学习模型中可以结合传统的统计方法，传统的统计方法增加模型复杂度也可以提升性能。

基于矩阵及向量的表示：14 页

核心思想：对于 m 个 d 维样本，将每个样本数据表示为行向量，数据矩阵 X 的行数即为样本个数。 $X: m \times d$ ，样本分为 C 类，真值矩阵 $Y: m \times C$ 。第一个隐藏层节点个数为 d_1 ，输出为 h_1 ，每个结点权值向量为行向量，长度为 d 。每个节点有一个偏置值 b 。对应的权值矩阵 $W_1: d_1 \times d$ ，偏置量 $b: 1 \times d_1$

求导：见习题课课件

二、深度学习的优化方法

高维空间的非凸优化问题：鞍点、局部最小值

1. 学习率调整：分段衰减、指数衰减
2. 梯度估计调整：动量法：利用前一次迭代的参数调整量，得到更为稳定的数值。
$$\Delta \theta_t = \rho \theta_{t-1} - \eta g_t$$
3. 梯度估计调整+自适应学习率：

Adam 算法： $m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t$ ， $v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2$ ，其中 $m_0=0, v_0=0$ ， $\hat{m}_t = m_t / (1 - \beta_1^t)$ ， $\hat{v}_t = v_t / (1 - \beta_2^t)$ ， $\theta_t = \theta_{t-1} - \eta * \hat{m}_t / \sqrt{\hat{v}_t + \epsilon}$

4. 设计有助于优化的模型：使用更多的线性函数；在各层之间增加跳跃连接；添加和网络中间隐含层相连的辅助目标函数。

5. 参数初始化：

- (1) 预训练初始化
- (2) 随机初始化：高斯分布初始化、均匀分布初始化
- (3) 固定值初始化（偏置量通常用 0 来初始化）

6. 深度学习的超参数：层数、每层神经元个数、学习率、正则化系数、mini-batch 大小。

7. 数据预处理

(1) 图像大小归一化

(2) 数值动态范围归一化：最小最大值归一化、标准化、PCA。梯度指向更接近目标的方向。

三、深度学习的正则化方法

通过正则化引入人工干预：

(1) 增加优化约束：L1 或 L2 约束、数据增强

(2) 干扰优化过程：权值衰减、Dropout、提前停止

1. L1 或 L2 正则化：见 35 页

2. 权值衰减：参数更新过程中引入一个衰减系数， $\theta_t = (1 - \lambda) \theta_{t-1} - \eta g_t$

在标准的随机梯度下降中，权值衰减和 L2 正则化的效果相同；在较为复杂的优化方法（如 Adam）中，权值衰减与 L2 正则化并不等价。

3. 丢弃法：在训练过程中每次迭代时，将各层的输出节点以一定概率随机置为 0。

4. 提前停止：使用一个验证集来测试每一次迭代的参数在验证集上是否最优。如果在验证集上的错误率不再下降，就停止迭代。

5. 数据增强：图像数据的增强主要是通过算法对图像进行转变，引入噪声等方法来增强数据的多样性。

第五节 卷积神经网络

一、卷积的定义

详见 pytorch 教材

二、卷积神经网络

(一) 相关定义

1. 架构：包含一个或多个卷积层、池化层（或称亚采样层、汇聚层）、全连接层

2. 目标函数：分类任务 softmax + 交叉熵

3. 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。

4. 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。

(二) 卷积层

1. 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。

2. 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）。

3. 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。

4. 卷积与全连接：将图像展开为向量，输入全连接层；保持图像空间结构，输入卷积层。

5. 卷积核尺寸：长、宽、输入通道数（与输入数据通道数相同，不额外设定）。

6. 其他常用超参数：移动步长、边界延拓、空洞卷积的膨胀率、分组数目。

7. 卷积层的参数：

(1) 张量表示：通常每个卷积层由卷积滤波器组构成，可表示成 4D 张量： $C_{out} \times C_{in} \times K_h \times K_w$ 。四个维度依次为输出通道数、输入通道数、卷积滤波器高度、卷积滤波器宽度。如果卷积核有偏置量，那么共有 C_{out} 个。详见 28 页。

(三) 池化层

1. 作用：尺度变化、参数降维。

2. 方式：取最大值、平均值、最小值。

3. 分类：最大值池化与平均值池化。

4. 卷积的扩展：

(1) 空洞卷积：在卷积模板中注入空洞，增大感受野尺寸。无空洞的卷积 dilation=1。

(2) 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。(收藏夹里有解释)

(3) 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。(收藏夹里有解释) 深度可分离卷积在做卷积运算时，只考虑区域，做完卷积再考虑通道。普通卷积操作同时考虑通道和区域。

三、卷积神经网络的训练方法

1. CNN 误差反向传播：见习题课课件。

2. BN(batch normalization)：逐层对数据进行尺度归一化，用于缓解梯度消失问题。对特征图按每一批次样本计算均值和方差，对数据各维度进行归一化处理，得到标准正态分布数据。

(1) 训练完成时：计算整个数据集上的均值 μ 与方差 σ ，用于测试过程。实践中，也可以用多个批次移动平均来计算。

(2) 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。

(3) 批量归一化可以看作一个特殊的网络层，加在每一层非线性激活函数之前。

作用：可以加速训练过程的收敛速度，避免梯度消失和梯度溢出。也可以看作一种正则化方法。

第六节 卷积神经网络的应用

(一) 图像分类

1. 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与任务上进行迁移学习。

(1) 预训练模型：随机初始化模型参数，在较大的数据集上从头训练，得到预训练模型（常用预训练模型具有网络下载链接）。

(2) 模型参数细调：用预训练模型参数进行模型初始化，根据目标域任务重设最后一个全连接层的输出节点数，有两种方式：

①整个卷积神经网络进行参数细调。

②卷积层相当于特征提取器，仅训练全连接层。

(二) 图像分割

1. 图像语义分割：传统图像分割是根据图像的颜色纹理进行划分同质区域，而图像语义分割是基于一个语义单元，例如将人、车等目标从图像中分割出来。利用全连接网络实现图像语义分割。

2. 图像的升采样：

(1) 重复元素

(2) 保留元素，其他位置赋 0

(3) 转置卷积：见收藏夹

(三) 目标检测

1. 传统流程：区域选择、特征提取、分类器分类

2. 深度学习——单阶段法

图像经过网络一次可以预测出所有的边界框。速度相对较快，适合移动平台计算。典型方法包括 SSD, YOLO 等。

3. 深度学习——两阶段法

先对输入图像生成候选框，然后再对每个候选框进行分类。例如 Faster R-CNN。

4. 目标检测思路

- (1) 可能的检测框过多——在有限集合上进行优选。
- (2) 分类方法未定义——结合分类任务进行训练。
- (3) 冗余的输出——基于物体实例的输出。
- (4) 前景背景样本数量不平衡——合理的采样、设计目标函数、多尺度处理等。

5. 交并比：测量两个框之间的相似性。0 表示不重叠，1 表示完全相同。给定两个集合 A、B，Jaccard 系数 $J = |A \cap B| / |A \cup B|$

6. 目标检测算法的评价

(1) 首先设定预测框与真值框 IoU 阈值，例如 0.5。若预测框与真值框的 IoU 大于阈值，且类别相同，则为 True Positive，否则为 False Positive。（注：每个真值框只允许有一个预测框为 True Positive）。由此可计算各类别的召回率和准确率。

实际应用中，采用不同的算法调整模式，得到不同的召回率和准确率。例如，调整目标检测算法参数，比如类别判定阈值；调整预测框与真值框 IoU 阈值。

- (2) 平均准确率 AP：对于单个类别，不同召回率下对应的准确率的平均。
- (3) mAP：多个类别 AP 的均值。
- (4) 运行速度 FPS：每秒处理图像数量。

7. 目标检测基础方法

(1) 给定锚点，选择多个不同尺度和宽高比的区域，称为锚框。预测每个锚框是否包含物体，如果是，则预测从锚框到实际边界框的偏移量，锚框加上偏移量为预测的外接框。利用预测偏移量与根据真值计算的真正偏移量的误差，构建训练过程中检测网络分支的目标函数。

(2) 锚框：每个锚框截取的图像都是一个分类任务样例。①预测每个锚框的类别②预测从锚框到真实边界框的偏移量。可能会生成大量的锚框，大部分是负类样本。

(3) 输出预测边界框：非极大值抑制（NMS）。

每个锚框生成一个边界框预测，一个物体会得到大量重叠的预测框。“非极大值抑制”是指选择得分最高的一个预测框作为输出，如果其他与之重叠的预测框 IoU 大于设定值，则删除。

8. 单阶段目标检测方法：SSD 与 YOLO

(1) SSD：在多尺度特征图上预测每个锚框的类别和边界框，对于指定均匀位置的像素，生成以像素为中心的多个锚框。

(2) YOLO：将输入图像均匀地切割成 $S \times S$ 锚框。

9. 两阶段目标检测方法：R-CNN

(1) 定义：在传统目标检测算法中，用 CNN 分别提取各个候选区域特征。

(2) Fast R-CNN 是在 CNN 输入的整个图像特征图上，采用传统方法提取候选区域，并利用 CNN 对候选区域进行预测。

(3) Faster R-CNN 利用基于 CNN 的区域生成网络 Region Proposal Network (RPN) 得到候选框。

①Conv layers

利用一组 conv+relu+pooling 层得到特征图，用于后续 RPN 层和全连接层。

②Region Proposal Networks (RPN)

RPN 网络用于生成区域候选框，通过 softmax 判断 anchors 是否属于某类物体或者背景，然后做区域候选框位置回归

③RoI Pooling (Region of interest pooling)

输入为特征图和区域候选框，由此截取 RoI 特征图，送入后续全连接层。

④Classification

利用 RoI 特征图判定类别，再次进行精细的区域候选框位置回归，然后利用 NMS 等后处理过程得到最终输出。

(4) Region Proposal Network (区域生成网络)

锚点 Anchor: 遍历特征图上的像素，设为锚点。

锚框 Anchor boxes: 在锚点处采用 K 个不同大小、宽高比的锚框

(5) Mask R-CNN: 在 Faster R-CNN 基础上，增加了一个用于预测区域候选框中目标前景掩膜的分支。

预测候选目标区域前景掩膜的分支为全卷积网络 (FCN)。利用区域预测网络输出的 RoI 候选框，截取特征图用于预测掩膜。

第七节 循环神经网络

一、传统循环神经网络 (见 pytorch 教材)

1. (1) Elman 网络: 将上一时刻隐含层状态作为隐含层的反馈输入。

(2) Jordan 网络: 将上一时刻网络输出作为隐含层的反馈输入。

2. 目标函数:

网络输出 k 维向量 y_t , $(y_t)_k$ 是 y_t 属于第 k 类的概率。样本真值标签 $g_t=i$, 则 $L_t=-\log(y_t)_i$, $L=\sum\{L_t\}$ 。

3. 误差随时间反向传播算法 BPTT: 见 19 页

4. 缺陷: 梯度消失/爆炸问题

二、门控机制循环神经网络

1. 长短时记忆单元 (LSTM)

LSTM 中引入了 3 个门, 即输入门、遗忘门和输出门, 以及与隐藏状态形状相同的记忆细胞, 从而记录额外的信息。LSTM 的输入为 $X_t(n \times d)$ 和 $H_t(n \times h)$, 输出为

$I_t = \text{sigmoid}(X_t W_{xi} + H_{t-1} W_{hi} + b_i)$, W_{xi} 是 $d \times h$ 维, W_{hi} 是 $h \times h$ 维, b_i 是 $1 \times h$ 维 F_t, O_t 同理。候选记忆细胞 $\hat{C}_t = \tanh(X_t W_{xc} + H_{t-1} W_{hc} + b_c)$, 参数维度同上。 $C_t = F_t \cdot C_{t-1} + I_t \cdot \hat{C}_t$, $H_t = O_t \cdot \tanh(C_t)$ 。

2. 门控循环单元 (GRU)

引入了重置门和更新门的概念。输入为 $X_t(n \times d)$ 和 $H_t(n \times h)$, 输出为 $R_t = \text{sigmoid}(X_t W_{xr} + H_{t-1} W_{hr} + b_r)$, W_{xr} 是 $d \times h$ 维, W_{hr} 是 $h \times h$ 维, b_r 是 $1 \times h$ 维, Z_t 同理。候选隐藏状态 $\hat{H}_t = \tanh(X_t W_{xh} + R_t \cdot H_{t-1} W_{hh} + b_h)$, 参数维度同上。重置门控制了上一时间步的隐藏状态如何流入当前时间步的候选隐藏状态。因此, 重置门可以用来丢弃与预测无关的历史信息。 $H_t = Z_t \cdot H_{t-1} + (1 - Z_t) \cdot \hat{H}_t$ 。

三、深层循环神经网络

1. 常用网络架构

(1) 双向长短时记忆网络及 CTC 解码

(2) 编码器-解码器

2. 双向循环神经网络

输入序列可按时间正向和反向分别送入两个层, 两层输入按对应时刻并接得到输出序列, 送入后续网络。

3. 连接时序分类 (CTC): RNN 中的解码层 (见 46-47 页及习题课课件)

(1) 在模式类别集合 Y 中, 增加一个空白标签 blank 符号。给定时间长度 T 的输入特征序列 x , 在时刻 t 网络输出 k 维向量 y_t , 每一维对应属于每一类的概率。在时刻 t 取概率最

大的类别标签 a_t 作为预测结果。预测的类别标签序列 a 的概率是各时刻预测的类别标签概率的乘积。输出为类别标签序列的概率是各种解码后可以得到 y 的序列的概率之和。

训练阶段的目标函数： $L = -\log(p)$

4. 案例：

基于 BiLSTM+CTC 的连写文字识别：预处理（图像归一化，二值化），输入层（图像→滑动窗序列），双向 LSTM 层（序列建模），Dropout 层（防止过拟合），Softmax 层（字符分类器），CTC 层（解码器）。

第八节 深度学习中的注意力机制

一、媒体信息处理中的注意力机制

1. 注意力系数：计算向量相似度：

加性模型： $v^T \tanh(Wx_i + Uq)$

点积模型： $x_i^T q$

缩放点积模型： $x_i^T q / \sqrt{d}$

双线性模型： $x_i^T W q$

2. 深度学习中的注意力机制：在特征表示中引入加权系数。

- （1）卷积神经网络：对多通道特征图进行加权。
- （2）循环神经网络：对特征序列进行加权。
- （3）基于自注意力机制的 Transformer：对特征序列进行加权。

二、卷积神经网络中的注意力机制

1. 卷积神经网络输出的特征图

- （1）基于空间位置的注意力机制
- （2）基于通道的注意力机制

2. 基于空间的注意力机制

- （1）Hard Attention：基于空间位置选取图像中一个合适的子区域，进行后续处理
- （2）Soft Attention：例如，空间位置掩膜作为注意力加权系数。

3. 基于通道的注意力机制——squeeze-and-excitation

- （1）Squeeze：权重预测分支对该输入特征图进行按通道的全局平均池化得到尺寸为 $1 \times 1 \times$ 通道数的向量。
- （2）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 $1 \times 1 \times$ 通道数的权重向量，分别乘以对应通道的特征图作为输出。

三、循环神经网络中的注意力机制

见 17 页

四、基于自注意力的 Transformer

1. 自注意力模型（收藏夹里有）

2. 多头自注意力

特征矩阵的通道拆分采用多头分支并行计算方式，并将各分支计算结果并接得到最终输出。

第九节 支持向量机

一、引言

方法	适用问题	模式类型	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		极大似然估计，极大后验概率估计	迭代学习
隐含马尔可夫模型	序列建模	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
支持向量机	两类分类	鉴别式	间隔最大化	序贯最小算法 (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

1. 支持向量机：一种基于统计学习理论的机器学习方法。

(1) 传统方法：主要研究渐进理论，即当样本趋于无穷多时的统计性质。

(2) 支持向量机：

①小样本条件下的统计学习方法。

②具备严格的数学理论基础和直观几何解释。

③处理不均匀、离散、稀疏的数据有明显优势。

④适用于样本有限情况下的优化求解。

(3) 判别函数： $G_i(x) > G_j(x)$ for all $j \neq i$

(4) 线性分类器： $y = w^T x + b$

样本 x 到超平面 H 的有向距离： $r = \frac{w^T x + b}{\|w\|}$

超平面的单位法向量： $n = \frac{w}{\|w\|}$

(4) K 近邻：如果一个测试样本在特征空间中的 k 个最相邻的标记样本多数属于某一类别，则该样本也属于这一类别。

(5) 最大间隔线性分类器：泛化能力强

目标：找到一个超平面，使得其能够尽可能地将两类数据正确分开，且分开的两类数据点距离分类界面最远。

一、优化问题建模

1. 问题分类：线性可分问题、近似线性可分问题、线性不可分问题。

(一) 线性可分问题

选取训练样本 (x_i, y_i) , $1 \leq i \leq n$, x_i 是向量, y_i 是标量。其中对于类别 $y_i=1$, $w^T x_i + b \geq 1$; 对于类别 $y_i=-1$, $w^T x_i + b \leq -1$ 。即 $y_i (w^T x_i + b) \geq 1$ 。目标是最大化间隔宽度为 $2/\|w\|$ 。最终求得 $w = \sum \{\alpha_j y_j x_j\}$, (x_j, y_j) 在支持向量集合中。

拉格朗日乘子法

(二) 近似线性可分问题

噪声和样本异常值的出现导致数据近似线性可分。引入松弛因子 m_i , $y_i (w^T x_i + b) \geq 1 - m_i$, $m_i \geq 0$, 目标是最小化 $0.5 \|w\|^2 + C \sum \{m_i\}$ 。参数 C 可防止过拟合。最终求得 $w = \sum \{\alpha_i y_i x_i\}$, $\sum \{\alpha_i y_i\} = 0$, $0 \leq \alpha_i \leq C$

坐标上升法：详见 28 页

序列最小优化算法：收藏夹

(三) 非线性问题

1. 特征空间：如果需要将样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。详见 33 页

2. 常用核函数：

线性核： $K(x_i, x_j) = x_i^T x_j$

多项式核： $K(x_i, x_j) = (1 + x_i^T x_j)^p$

高斯核： $\exp(-\|x_i - x_j\|^2 / (2 \sigma^2))$

Sigmoid: $K(x_i, x_j) = \tanh(\beta x_i^T x_j + \beta_i)^p$

最终求得 $w = \sum \{\alpha_i y_i \phi(x_i)\}$

第十节 统计模式识别

4 页有一个分类表值得一看。

一、贝叶斯决策

1. 相关定义见 6-7 页

2. 统计学中的概率学派与贝叶斯学派

(1) 相同点：最大似然函数在概率学派和贝叶斯学派都有重要作用。其思想是认为已观测数据的概率分布是最大概率，最大概率对应的模型就是需要找的模型。

(2) 不同点：概率学派认为模型参数是常数，二贝叶斯学派认为模型参数是变量，用概率去描述模型参数的不确定性。

3. 模式识别中的贝叶斯定理：9 页

后验概率 = 先验概率 \times 类条件概率密度函数 / 样本出现概率

4. 贝叶斯决策：13 页

贝叶斯决策具有最小错误率。

二、正态分布下的贝叶斯决策

1. 相关定义：见 16 页

2. 最大似然估计：

$$l(\theta) = \ln p(D | \theta) = \sum \{p(x_i | \theta)\} \quad \hat{\theta} = \arg \max \{l(\theta)\}$$

$$\hat{\mu} = \hat{\theta}_1 = \sum \{x_k\} / n \quad \hat{\Sigma} = \hat{\theta}_2 = \sum \{(x_k - \hat{\mu})(x_k - \hat{\mu})^T\} / n$$

$$\text{无偏估计: } \hat{\Sigma} = \hat{\theta}_2 = \sum \{(x_k - \hat{\mu})(x_k - \hat{\mu})^T\} / (n-1)$$

假设类条件密度概率函数为正态分布, $p(x | w_i) = \exp\{-0.5(x - \mu_i)^T \Sigma_i^{-1}(x -$

$u_i\} / ((2\pi)^{0.5d} |\Sigma_i|^{0.5})$ 。判别函数定义为: $G(x_i) = p(x|w_i)p(w_i) = p(w_i) \exp\{-0.5(x-u)^T \Sigma_i^{-1}(x-u)\} / ((2\pi)^{0.5d} |\Sigma_i|^{0.5})$ 。取对数得 $g(x_i) = -0.5(x-u)^T \Sigma_i^{-1}(x-u) - 0.5d \ln(2\pi) - 0.5 \ln |\Sigma_i| + \ln p(w_i)$

3. 正态分布协方差矩阵的不同情形 (假设各类先验概率相等)

(1) 协方差矩阵=单位阵 \times 方差, 是最小欧氏距离分类器, 也就是线性分类器。见 22 页。

(2) 各类协方差阵的元素都相等, 是最小马氏距离分类器。(21 页)

线性判别函数(LDF): $g(x) = A_i^T x + b_i$, $A_i = \Sigma^{-1} u_i$, $b_i = -0.5 u_i^T \Sigma^{-1} u_i$

二次判别函数(QDF): $g(x) = (x-u)^T \Sigma^{-1} (x-u) + \ln |\Sigma| = \sum\{(\phi_i^T (x-u))^2 / \lambda_i\} + \sum\{\ln(\lambda_i)\}$

MQDF 类型一: 在协方差矩阵的对角线上加上一个小的常量 h^2 , 保证矩阵的满秩 (非奇异性), 由此取代协方差矩阵的最大似然估计, 即为 $\Sigma_p = \Sigma + h^2 * I$, $g(x) = \sum\{(\phi_i^T (x-u))^2 / (\lambda_i + h^2)\} + \sum\{\ln(\lambda_i + h^2)\}$ 。

MQDF 类型二: 用 h^2 代替较小特征值, 也就是说, 只有前 k 个特征值有效。 h^2 有多种取值方式, 比如取所有较小特征值的平均值。

三、混合高斯模型

(一) 聚类

1. 基于误差平方和准则的聚类: 46 页

k-means 聚类方法:

(1) 输入: 聚类中心数目 k 和 n 个样本。

(2) 输出: k 个聚类中心, 使平方误差准则最小。

(3) 算法步骤:

①为每个聚类确定一个初始聚类中心, 这样就有 k 个初始聚类中心

②将样本集中的样本按照最小距离原则分配到最邻近聚类。

③使用每个聚类中的样本均值作为新的聚类中心。

④重复②、③直到聚类中心不再变化。

⑤结束, 得到 k 个聚类中心。

(4) 性能分析:

①简单、快速

②当数据聚类是紧致的, 而类簇之间区别明显时, 效果较好。

③在类的平均值被定义的情况下才能使用。

④必须事先给出 k 。

⑤对于不同的初始值, 可能会给出不同的结果。

⑤对于噪声和孤立点数据较为敏感。

(5) 改进方法

①多设置一些不同的初始值, 但比较费时和浪费资源。

②ISODATA 算法: 通过类的自动合并和分裂, 得到较为合理的类型数目 K 。

(二) 混合分布的概率密度估计问题

1. 问题框架

(1) 所有样本都来自于 K 种类别, K 已知。

(2) 每种类别的先验概率未知。

(3) 类条件概率的数学形式已知, 但具体参数未知。

(4) 样本类别未被标记。

2. 混合高斯模型的参数估计:

$$p(x|w) = \sum \{ \pi_k N(x|u_k, \Sigma_k) \}, \quad \sum \{ \pi_k \} = 1.$$

3. 期望-最大化 EM 算法:

给定一些观察数据 x , 假设 x 符合如下混合高斯分布:

$$p(x) = \sum \{ \pi_k N(x|\theta_k) \}, \quad \text{求混合高斯分布的参数 } \pi_k, \theta_k$$

最大似然估计:

$$u_k = \sum \{ x_k \} / N$$

$$\Sigma_k = \sum \{ (x_k - u_k)(x_k - u_k)^T \} / N_k$$

$$\pi_k = N_k / N$$

算法过程:

- (1) 初始化 K 个高斯分布参数 u_k, Σ_k , 初始化 π_k 并保证 $\sum \{ \pi_k \} = 1$ 。
- (2) 根据目前的高斯分布参数, 对样本 x 的类别隐藏变量 Z_{nk} 求期望, $\gamma(z_{nk})$ 表示 x_n 属于第 k 类的概率。 $\gamma(z_{nk}) = \pi_k N(x_n|u_k, \Sigma_k) / \sum \{ \pi_j N(x_n|u_j, \Sigma_j) \}$
- (3) 对高斯分布参数求最大似然估计, $u_k^{\text{new}} = \sum \{ \gamma(z_{nk}) x_n \} / N_k$, $\Sigma_k^{\text{new}} = \sum \{ \gamma(z_{nk}) (x_n - u_k^{\text{new}})(x_n - u_k^{\text{new}})^T \} / N_k$, $N_k = \sum \{ \gamma(z_{nk}) \}$, $\pi_k^{\text{new}} = N_k / N$,
- (4) 迭代计算第 2, 3 步, 直到满足参数收敛条件或停止条件。

第十一节 隐含马尔可夫模型

一、定义

1. 马尔可夫过程: 模型“将来”的条件分布仅受“现在”影响, 不受“过去”影响。
2. 马尔可夫链: 参数集与状态空间都是离散的马尔可夫过程称为马尔可夫链。
3. 状态转移概率矩阵: 每一行代表一个当前状态, 行中不同的列表示从当前状态转移下一状态的概率。
4. 基本原理: 11 页。
5. 隐含马尔可夫模型
状态序列是马尔可夫链, 用转移概率描述; 每一状态对应一个可以观察的事件, 用观测概率描述。
6. 描述隐含马尔可夫模型的三元组: $\lambda = (\pi, A, B)$ 。
 A 是与时间无关的状态转移概率分布, B 是给定条件下观察值概率分布, π 是初始状态空间概率分布。(15 页)
7. 基本假设 (16 页)
 - (1) 状态序列的马尔可夫性
 - (2) 齐次性, 即状态转移概率与具体时刻无关。
 - (3) 观测序列之间的独立性
8. 训练与求解
 - (1) 直接计算, 复杂度 $O(TN^I)$
 - (2) 引入前向、后向辅助变量: $O(N^2T)$见 20-29 页

第十二节 特征提取与降维

一、特征提取

1. 定义: 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。

2. 一维信号时频分析 (7 页)

(1) 傅里叶变换

(2) 短时傅里叶变换 (Gabor 变换)

$STFT(\tau, \omega) = \int s(t) f(t - \tau) e^{-j\omega t} dt$ 。s(t) 为信号, g(t) 为窗函数。

(3) 小波变换

$W_s(a, b) = \int s(t) \phi((t-b)/a) / \sqrt{a}$

3. 二维 Gabor 小波函数

$f(x, y) = \exp\{-[(x-x_0)/\alpha]^2 + [(y-y_0)/\beta]^2\} * \exp-i\{u_0(x-x_0) + v_0(y-y_0)\}$

$w_0 = \sqrt{u_0^2 + v_0^2}$, $\theta_0 = \arctan(v_0/u_0)$ 。(x₀, y₀) 指定图像中的像素位置, (α, β) 为有效的宽度和高度, (u₀, v₀) 为调制参数。

二、特征降维

1. 目的: 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。

2. 特征降维由原始数据生成最有效的特征, 其性质包括①同类样本的不变性②异类样本的鉴别性③对噪声的鲁棒性

3. 常见方法: 特征选择、特征变换

4. 特征选择:

(1) 直接选择法: 从 n 个原始特征中选出用于分类的 d 个特征, 使可分性判据 J 的值满足最大 (18 页)

(2) 变换法: 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。

①主成分分析 (PCA): 最小均方误差准则下保留原始数据信息。无监督, 取数据投影方差最大的方向。

②线性判别分析 (LDA): 最小均方误差准则下区分多类数据。有监督, 取分类性能最好的投影方向。

假设有 n 个 p 维样本 x_i, 标签为 y_i, 对应 C 个类别 w_j, 每一类样本数量为 N_j, 均值为 u_j, 样本总体均值为 u。类内散度矩阵 $S_w = \sum \{ \sum \{ (x_{ki} - u_i)(x_{ki} - u_i)^T / N \}$ 表示样本点 x 围绕各类均值的散布情况。类间散度矩阵 $S_b = \sum \{ n_i (u_i - u)(u_i - u)^T / N \}$ 总散度矩阵 $S_T = S_w + S_b$, $S_T = \sum \{ (x_i - u)(x_i - u)^T / N \}$ 。对 $S_w - 1 S_b$ 做广义特征值分解, 得到的前几个特征值、特征向量用于降维。

(一) 主成分分析表示各类中心围绕均值的散布情况。

20-34 页

重点: 利用 PCA 进行特征降维

(1) PCA 优点:

①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。

②在最小均方误差准则下, 是最佳变换。

(2) PCA 缺点: 变换矩阵随样本数据而异, 无快速算法。

第十三节 认知的生物机制

一、概述

1. 对人的认知的研究层次:

(1) 分子层次: 脑中最基本的成份——各种分子的功能 (神经递质、蛋白质等)。

(2) 细胞层次: 神经元是如何行使其功能。

(3) 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。

(4) 认知层次：神经系统是如何产生认知和协调的行为？

2. 通过技术手段研究人的认知：

(1) 脑电图 (EEG)：通过电极记录下来的脑细胞群的自发性、节律性电活动。①皮层电图：在病人进行脑外科手术时，直接在皮层表面引导的电位变化。②脑诱发电位：在感觉传入冲动的激发下，脑的某一区域可以产生较为局限的电位变化。

(2) 脑成像技术：(1) 正电子发射层描记术 (PET)：使用对比剂，放射性同位素。(2) 事件相关电位 (ERP)，脑磁图 (MEG)：高时间分辨率，低空间分辨率。(3) 磁共振成像

(MRI)：非侵入性，没有辐射，应用广泛。

3. 磁共振成像过程：人体未进入静磁场时，体内氢原子群磁矩自然无规律排列。进入静磁场，所有自旋的氢原子重新排列定向，磁矩指向 N 或 S 极。通过射频线圈与静磁场垂直方向施加射频脉冲，受检部位氢原子吸收能量并向一个方向偏转与自旋。射频脉冲停止，核磁弛豫开始，氢原子释放吸收的能量重新回到原来自旋的方向。释放的电磁能转化为磁共振信号，经梯度磁场做层面选择和相位编码及频率编码，经傅里叶变换和计算机处理形成图像。

(3) 反应时

①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，同时也可以把刺激理解为信息。

②行为：行为是有机体对于所处情境的反应形式，可分解为刺激、生物体、反应三项因素。

③心理学实验：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。

④反应时间 (RT)：从刺激的呈现到反应的开始之间的时间间隔。在反应的潜伏期中包含着感觉器官、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：人在辨识字体颜色时会受到字义的影响。

二、感觉与知觉

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

2. 知觉：将感觉信息组合成有意义的对象，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。

3. 视觉感知：光刺激于人眼所产生。

(1) 视觉的生理机制：折光机制、感觉机制、传导机制、中枢机制。眼球包括眼球壁和眼球内容物。眼球壁分为三层：外层为巩膜和角膜（屈光作用）；中层为虹膜、睫状肌和脉络膜；内层是视网膜（感光）和视神经内段。眼球内容物包括晶体、房水和玻璃体，都是屈光介质。

(2) 视网膜是眼球的光敏感层。①外层的锥体细胞和棒体细胞：棒体细胞主要感受物体的明暗，锥体细胞主要感受物体的细节和颜色。②中间的双极细胞③内层的神经节细胞

4. 视觉的传导机制：电信号从感受器产生以后，沿着视神经传至大脑。传递机制有三级神经元实现 (1) 视网膜双极细胞：具有侧抑制作用。(2) 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。(3) 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。

5. 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。

6. 视觉的中枢机制：人类的视觉皮层包括初级视皮层（V1，也称纹状皮层）和纹外皮层（V2, V3, V4, V5 等）。

7. 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。

（1）V1 视神经细胞主要有三种感受野：同心圆感受野，简单感受野，复杂感受野。

（2）同心圆感受野（中心-周边感受野）：

①给光（on 型）：感受野中心光兴奋，周边光抑制。②激光（off 型）：反之。

（3）简单感受野：对大面积弥散光刺激没有反应，而对有一定方向或朝向的条纹刺激有强烈反应。若该刺激物的方向偏离该细胞“偏爱”的最优方位，则细胞反应停止或骤减。同时，它们对该类视觉刺激的位置和空间频率也表现出了明显的选择性。（简单细胞，比较适合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。）

（4）复杂感受野：对一定方向或朝向的条纹刺激有强烈反应，但是对于视觉刺激在视野中的位置没有选择性。即满足方向偏好的视觉刺激，无论出现在感受野的任何位置，都能激发视神经细胞的响应。对应于特定方向的条形刺激，具有位置不变性。

8. 不同类型感受野之间的关系

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。

9. 同心圆感受野的数学模型：

$$\text{DoG}(r) = A \exp(-r^2/a^2) - B \exp(-r^2/b^2)$$

R 为感受野中一点与中心点的距离，a、b 表示标准差。A>B, a<b 为 on-中心，反之对应 off 中心。

10. Gabor 滤波器与人类视觉系统中简单细胞的视觉刺激响应类似。

11. 纹外皮层：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。对波长（颜色）有选择性的细胞集中在细条纹中，对运动方向性有选择性的细胞存在于粗条纹中，对形状敏感的细胞则在粗条纹和细条纹中都有所分布。

11. 知觉的信息加工过程

（1）自下而上加工（数据驱动加工）：指由外部刺激开始的加工，通常是先通过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连续加工，而达到对感觉刺激的解释。

自下而上加工从视觉输入开始，以视觉的再现结束。视觉输入传递顺序分为高低阶段，低阶段依赖视觉输入，高阶段产生知觉再现。低阶段的结果不受高阶段影响。

（2）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，从而调整特征器，引导对细节的注意等。也就是说，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。

三、注意

1. 定义：人的心理活动对一定对象的指向和集中。

2. 功能：信号检测、选择性注意、分配性注意。

3. 特征：选择性、持续性、注意的转移。

4. 注意的选择性：

（1）指向性：是指人在每一瞬间，其心理活动或意识选择了某个对象，而忽略了另一些对

象。

(2) 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。

5. 注意选择的理论模型

(1) 过滤器模型：来自外界的信息是大量的，但人的神经系统高级中枢的加工能力是有限的。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。

①感觉登记：瞬时记忆又称感觉记忆或感觉登记，内容是最初刺激的精确表象。

②选择性过滤：过滤器决定哪些刺激接受进一步加工。只有通过选择性滤波器的刺激才能被“知晓”。

③觉察器

④短时记忆

(2) 衰减模型（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。过滤器允许一个通道的信息通过，也允许另一个通道的信息在受到衰减、强度减弱后通过，但其中一些信息仍可得到高级加工。

(3) 过滤器模型与衰减模型比较：

①区别：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。过滤器模型中的过滤器是“全或无”性质，未选择的通道是完全关闭的；而衰减器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。

②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，需要过滤器调节。过滤器的位置都处于初级分析和高级分析之间。过滤器的作业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。

(4) 反应选择模型

几个输入通道的信息均可以进入高级分析水平，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。

感觉登记——知觉分析——反应选择——组织输出。

(5) 知觉选择模型与反应选择模型比较

刺激——觉察——（知觉选择）——识别——（反应选择）——复述与反应择定——反应

6. 注意的认知资源分配：

(1) 双加工理论：

①控制性加工：受到人的意识控制与认知资源的限制，需要意识的加工。其容量有限，可灵活用于变化着的环境。习得后加工过程很难改变。经过大量的练习后，可能转变为自动加工。

②自动加工：不受人控制的加工，也不受认知资源的限制，无需应用注意。没有容量限制，一旦形成就很难改变。

7. 注意的生理机制

(1) 朝向反射：情景的新异性引起一种复杂而又特殊的反射，是注意的最初生理机制。

(2) 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。

(3) 大脑皮层：产生注意的最高部位。

8. 计算机视觉中的注意力机制：将内部经验与外部感觉对齐从而增加部分区域的观察精细度的机制。

四、记忆

1. 定义：记忆是在头脑中积累和保存个体经验的心理过程。/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。

2 记忆的子系统：感觉记忆（SM），短时记忆（STM），长时记忆（LTM）。

3. 记忆的生理基础

（1）前额叶：短时记忆

（2）额叶：语义与情节记忆

（3）颞叶：参与长时语义与情节记忆的整合与存储，对短时记忆单元中的新材料加工也起作用。

（4）杏仁核：新情绪记忆信息的整合。

（5）海马：整合新的长时语义和情节记忆。

（6）小脑：程序性记忆。

（7）皮层运动区：程序性记忆。

4. 感觉记忆

（1）定义：感觉记忆又可称为瞬时记忆，是一种信息储存时间以毫秒或秒计的记忆。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。

（2）特点

①存储时间非常短：图像记忆在几百毫秒内，声像记忆可达 4 秒。信息加工只是初步的

（但可以进行信息整合）。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。

②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。

③记忆过程是无意识的自动化的，人无法控制。

5. 短时记忆：一种信息存储时间为一分钟以内（约 15-30 秒）的记忆。又可被称为电话号码式记忆。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。

（1）记忆的组织是一种表列等级结构，从短时记忆向长时记忆存入一项需要 5-10 秒。

（2）短时记忆信息的编码：在记忆系统对信息进行转换，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。

（3）短时记忆信息的存储、遗忘：短时记忆的保持时间为 15-30 秒，内容会随时间逐渐减少。

（4）影响因素及遗忘原因：

①干扰作业难度大小，记忆材料的熟悉程度。

②痕迹衰退说：记忆痕迹将随时间而消退。

③干扰说：已有信息的干扰。

6. 长时记忆：将信息保持存储时间在一分钟以上的记忆。存储信息的数量随时间的推移而逐渐下降。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，例如扭曲和错觉。

68-69 页有图。

五、思维

1. 定义：思维是借助语言、表象或动作实现的、对客观事物的概括和间接的认识，是认识的高级形式。

2. 特征

（1）概括性：在大量感性材料的基础上，把一类事物共同的特征和规律抽取出来，加以概

括。

(2) 间接性：借助于一定的媒介和知识经验对客观事物进行间接的认识。

(3) 对经验的改组：思维是探索 and 发现新事物的过程，需要对已有的知识经验不断进行更新和改组。

3. 过程：分析与综合，比较，抽象与概括

4. 种类：直观动作思维和形象思维和逻辑思维，经验思维和理论思维直觉思维和分析思维，辐合思维和发散思维，常规思维和创造思维。

5. 人的思维方式：心理学家认为，我们大脑中存在两套思维系统。

(1) 无意识运作，快速的：情绪，本能反应。(2) 受控制运行，通常与行为、选择和专注等主观体验相关：理智思考。

6. 批判性思维与创新：

(1) 批判性思维：批判性思维是面对认识的对象，做出肯定什么、否定什么、或要有什么新见解、新举措的一个系列思考过程。要得出合理的结论必须有正确的思考方法和途径。

(2) 学术研究中的批判性思维：充分了解研究对象，做到全面、客观、准确。不迷信已有的结论，更要善于提出问题、分析问题。不断反思自己的思维模式。自我批判已经形成的思维定式和习惯的研究方法，从新角度以新方式来思考问题。超越自我是创新的必经阶段。

第十四节 媒体与认知相互作用

一、类脑认知计算

1. 文字识别中的字符形变问题

(1) 场景文本：变形尺度大，存在大量的二维全局形变，如倾斜、弯曲。

(2) 手写文本：变形复杂多样，微小形变和全局形变共存。

2. 像素级图像修正方法：

$I'(x, y) = m(x, y) \cdot I(x + \Delta x, y + \Delta y)$ 。 I' 为校正图像， m 为采样幅度值， I 为原始图像， Δx ， Δy 为采样偏移量

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训练。

3. RNN 中的相邻混合输出方法

(1) RNN 基本公式：

$h_t = \text{RNN}(h_{t-1}, x_t)$

$h_t' = h_t + \lambda_t' \cdot h_{t-1}$

(2) 相邻输出混合方法：将相邻的隐藏状态线性组合，组合系数 λ_t' 由 beta 分布随机采样得到。

(3) shortcut 连接

引入相邻隐藏状态的随机线性组合。目标函数对隐藏状态 h_t' 的梯度能够反向传播值隐藏状态 h_t 和 h_{t-1} ，有利于网络的优化。

4. 信息的获取与利用

(1) 模式类概率空间 Y 上的系统熵：

$H(Y) = -\sum \{P(w_i) \log P(w_i)\}$ 。

$H(Y)$ 描述了样本模式类信息源的不确定性，也表示识别过程需要减少的信息熵

$H(Y) \leq \log(n)$ ，在各模式先验概率相同时取等号。

(2) 从信息熵的角度理解识别过程

特征熵 $H(X)$ 表示样本的随机特征变量在样本特征概率空间中包含的特征信息容量。

条件特征熵 $H(X|Y)$ 表示样本模式类别确定后，样本特征的平均信息量，它说明了样本的特征和样本的模式类别之间的相关程度。

(3) 模式识别信息熵

模式类概率空间 Y 与样本特征概率空间 X 之间的互信息 $I(X, Y) = H(x) - H(X|Y)$ 。互信息表示模式类概率空间 Y 和样本特征概率空间 X 所共有的信息量。

(4) 有关模式识别信息熵的定理

模式识别的学习过程是样本特征不确定度减小的过程。一信息熵表示，样本特征不确定度从样本特征熵减少到样本条件特征熵 $H(X|Y)$ 。学习过程的熵减称为学习熵减 $\Delta H_L = I(X, Y) = H(x) - H(X|Y)$ 。

5. 学习与识别过程的熵变化

(1) 学习过程：

互信息是决定模式识别性能的信息量。 $\Delta H_L = I(X, Y) = H(x) - H(X|Y)$ 。

(2) 识别过程：

后验熵反映识别不确定度。 $H(Y|X) + H(Y) - I(Y|X)$

二、信息的获取与利用

1. (1) 被动式获取：视觉，光场相机（包含一个主镜头，一个微透镜阵列和一个数字图像传感器）。

(2) 主动式获取：超声探测，结构光成像

2. 获取光场的手段：

(1) 微透镜阵列：在普通成像技术的镜面处加入一个微透镜阵列，每个微透镜记录的是光线对应相同位置不同视角的场景图像。

(2) 相机阵列：通过相机在空间的一定分布来同时抓取一系列视角略有差别的图像。

(3) 光场获取前沿技术：用多层透明图像传感层取代微透镜。

3. 深度信息获取：

(1) 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。

(2) TOF：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。

(3) 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。

三、创造新媒体

1. 视觉暂留与电影：

(1) 人眼观看物体时，成像于视网膜上，并由视神经输入人脑，从而感觉到物体的像。但当物体移去时，视神经对物体的印象不会立即消失，而要延续 0.1-0.4 秒的时间。由于视神经的反应速度，光作用于视网膜所引起的视觉，在光停止作用后，仍保留一段时间，其时值约为 1/16 秒，对于不同频率的光有不同的暂留时间。

(2) 电影胶片以每秒 24 格画面匀速转动，一系列静态画面由于视觉暂留作用形成连续的视觉印象。

(3) 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。

2. 立体视觉与立体显示

(1) 原理：人们通过生理层面的 4 种暗示和心理学层面的 6 种暗示。三维显示技术所要解决的核心问题是如何为人眼感知三维物体提供所需的上述十种深度暗示信息。

(2) 基于心理深度暗示形成 3D 效果图像

- ①线性透视：根据人们的观察习惯，景物向远处延伸时，所观察到的尺寸逐渐缩小。
- ②像的大小：很多物体的实际尺寸在人脑中都有一个固有的先验知识，因此可以通过图像的大小感知物体的远近，从而提供一种心理暗示。
- ③重叠：两个物体轮廓的重叠关系会产生暗示，通常认为被遮挡的部分处在下方或者远处等。
- ④光照及阴影：二者会产生一种极强的深度暗示。
- ⑤结构梯度：与线性透视类似，当我们注视均匀梯度是，其表面粗糙度方面的梯度会产生一种深度暗示。
- ⑥面积透视：在观看一幅二维图像时，人们总认为看起来比较模糊的景物处在远方，这是因为在人们的实际生活中，远处景物发出的光线在传播中被空气中的微粒散射而显得模糊。

（2）基于生理深度暗示形成 3D 效果图像

- ①调节：通过睫状肌中的平滑肌调节眼睛晶状体的焦距，使观察者可以看清楚远近不同的景物。
- ②汇聚：当观察者的眼部肌肉被拉伸使眼球略微转向内侧以便对着三维物体上某一点观看时，两只眼睛的视轴所组成的角度称为会聚角。通常来说，物体离观察者越近，则会聚角越大，物体远则会聚角小。
- ③双目视差：双目视差是由人眼的瞳孔间距（平均 6.5cm）所引起的。对于同一景物，左右眼看到的是有差异的图像。
- ④移动视差：如果观察者的观察位置发生变化，观察到的三维物体也会相应地发生变化，这个效应称为移动视差。

基于双目视差的三维显示技术是目前市场上的主流技术

（3）真三维显示技术

真三维显示技术所要解决的核心问题是如何为人眼感知三维物体提供多种深度暗示信息。三维显示技术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。进一步地，如果某一种三维显示技术还能同时提供心理学上的 6 种深度暗示，那么它将更符合人眼的观察习惯。真三维显示技术不会造成观看者的视觉疲劳，其显示的图像更加真实。

（4）体三维显示技术：一种基于多种深度暗示的真三维显示技术。通过特殊方式来激励位于透明显示空间内的物质，利用光的产生、吸收或散射形成体素，并由许多分散体素构成三维图像。或采用二维显示屏旋转或层叠而形成三维图像。

- ①基于动态屏的体三维显示：依靠机械装置旋转或移动平面显示屏，利用人眼的视觉暂留效应实现空间立体显示效果。
- ②基于上转换发光的体三维显示：利用两束不同波长的不可见光束来扫描和激励位于透明体积内的光学活性介质，在光束的交汇处取得双频两步上转换效应而产生可见光荧光，从而实现空间三维图像的显示。
- ③基于层屏的体三维显示：使用高速投影机将待显示物体的深度截面连续投射到与显示体相对应的深度位置上，且保证在较短时间（如 1/24s）内完成在显示体上的投影成像。利用人眼的视觉暂留效应，可在显示体前方任意位置观看。

3. 虚拟现实

（1）虚拟现实 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。

科学问题：如何把抽象数据空间表示为直观物体，如何利用多通道感应获得信息，如何使人与机器高度和谐交互。

(2) 虚拟现实待研究的问题:

①现实中需要再现的场景: 已经发生过的历史场景, 预想发生的事件演示, 科学幻想场景, 军事演练与工程预演, 虚拟人生等。

②科学研究需要可视化、可交互的 3D 空间:

物体内部 3D 探测数据, 人体检测数据, 流体动力数据, 微观结构数据等。

③现代信息社会需要自然人机交互方式: 键盘鼠标已不能适应三维交互。普适计算与便携式交互、自然操作中的行为检测等。

(3) 虚拟现实的思想:

人的感知系统接收虚拟真实场景→临场感。特征: 人所看到的场景随视点的变化而实时变化。

虚拟现实技术的三个特征: 沉浸、交互、幻想。

(4) 沉浸: 通过多种技术手段使用户置身于一个多维信息空间。分为视觉沉浸、听觉沉浸、触觉沉浸、嗅觉沉浸、味觉沉浸、辐射热感等。

技术难点: 感知与 VR 系统的交互需要。通过各种传感器, 感受虚拟现实环境。

视觉沉浸的两种实现方式: ①窗口显示法: 主要是通过显示屏来观察计算机生成的虚拟环境。缺点: 无法在虚拟空间漫游。②头盔显示法: 利用头盔隔离真实世界, 提供视觉虚拟环境。通过左右眼视差图像, 产生真实立体感, 操作者可获得沉浸于虚拟现实的感觉。

触觉沉浸: 使操作者可以操作并感触虚拟空间中的虚拟物体。最常用的触觉沉浸工具是数据手套。

(5) 交互: 交互性是指操作者对虚拟环境中的物体的可操作程度, 和从环境中得到的反馈的自然程度。主要借助于各种专用交互设备(立体显示器, 数据手套, SPIDAR 等), 使操作者如同在真实环境中一样, 与虚拟环境进行交互。

(6) 构想: 设计和创建 VR 新系统

探索 VR 机理, 加深对客观世界事物理解。VR 环境构想可以是远古环境、现实中的未来环境、微观环境、艰险环境等。提供体验: 使操作者获得如同亲历的震撼效果。飞行驾驶: 处理正常情况中难以遇到的复杂问题。手术导引: 体验前人长期积累的知识和经验。

(7) 虚拟现实系统构成: VR 是一个由高性能硬件, 软件和多种传感器组成的交互系统, 由 5 个关键部分构成。①虚拟环境数据: 包含三维模型, 环境定义数据库。②驱动软件: 提供观察和参与虚拟环境的功能。③计算机: 系统运算核心。④输入装置: 用于采集信息, 产生随动图像, 如三维鼠标, 数据手套。⑤显示虚拟环境视图, 如头盔, CAVE

4. 增强现实 (AR): 在虚拟现实的基础上发展起来的新技术, 也被称为混合现实。通过计算机系统提供的信息增加用户对现实世界感知。将计算机生成的虚拟物体、场景或系统提示信息叠加到真实场景中, 从而实现对现实的增强。

AR 通常首先借助于某种设备(最典型的是摄像头)获取“现实”的影响, 与传统视频应用直接显示影像不同, 再通过信息技术的处理, 在原生的影像上叠加文字、声音、虚拟图像图形信息之后再展示给用户。

应用领域: 虚拟现实和增强现实技术, 为我们提供实际场景不能实现的媒体环境, 通过与该环境的相互作用, 可以为我们带来新的认知体验。

媒体与认知的相互作用: 以人为本。认知系统通过媒体获得信息, 进一步获取知识; 媒体对认知系统的拓展: 媒体技术为提高人类认知客观世界的能力提供有效的手段和工具。创造新的媒体: 研究认知机理, 通过媒体信息智能管理, 创造新的媒体形式。新媒体——信息新的表现形式。