

《人工智能导论》习题及参考答案（后四次作业）

第五次作业

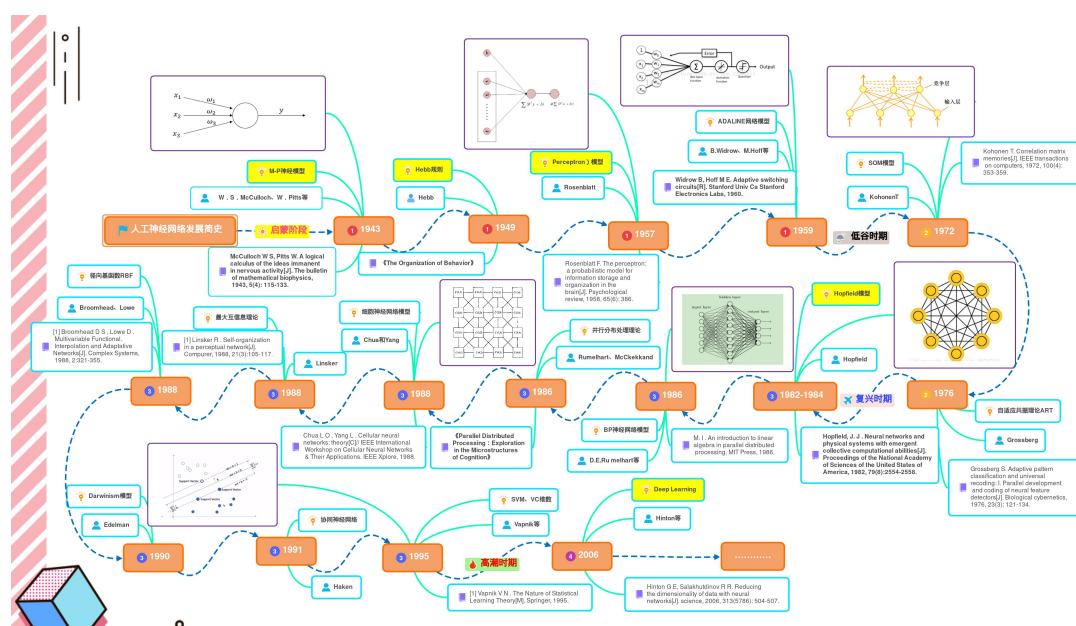
题目 1

画出一条人工神经网络的发展脉络图（截止时间 2010 年），其上标注你认为具有里程碑意义的方法，并注明时间、作者、参考文献。

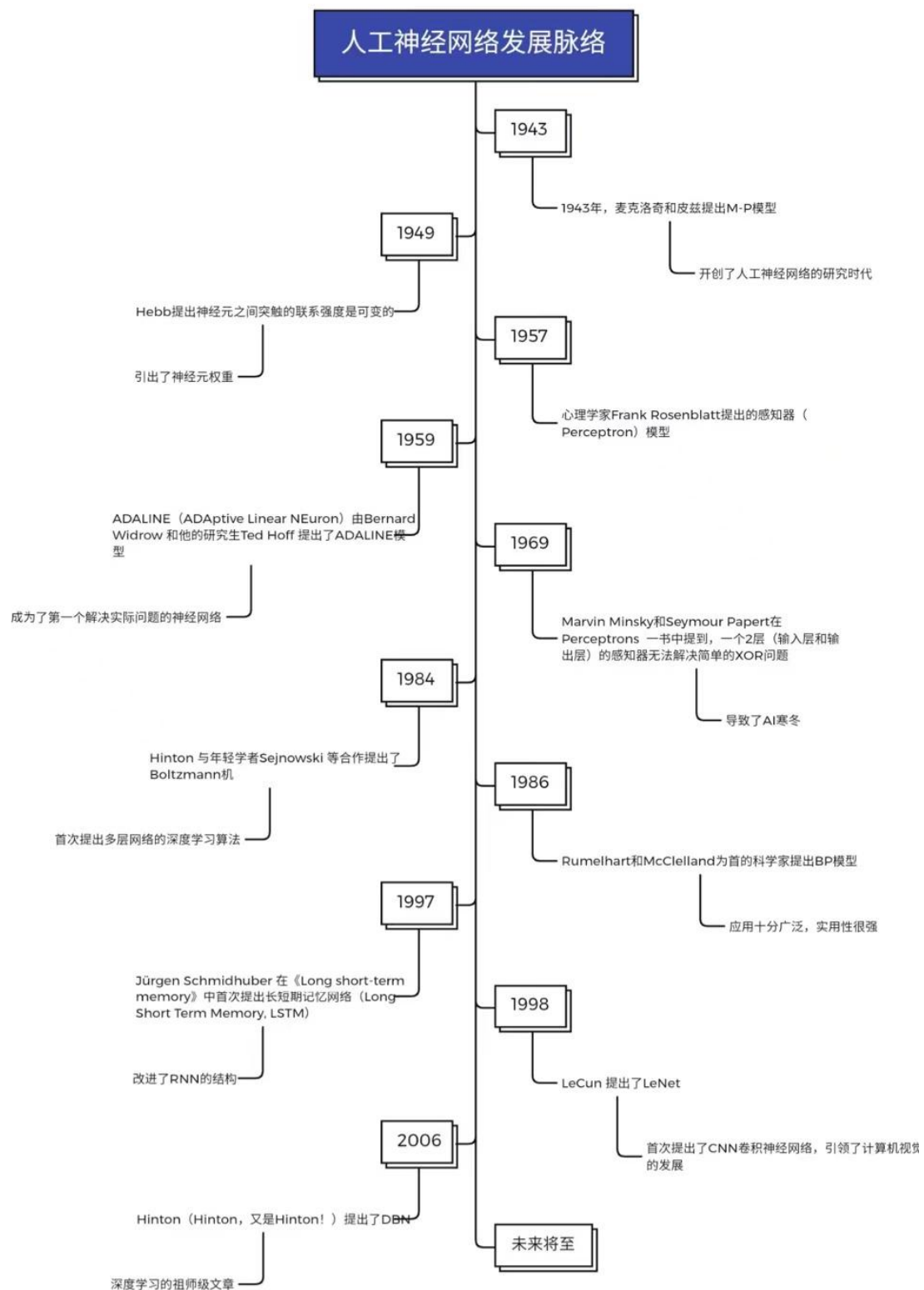
每项方法用一句话总结或评价：你为什么认为它是一项具有里程碑意义的工作//或它的哪项特点/进展对后面的工作具有重要的启发/奠基作用？

- 不限于本课件中使用或提及的插图、方法
- 本课件标注的时间、作者未必准确无误
- 本课件内容仅供参考，请自行检索有用信息或文献

参考答案：



*张永生同学



*陈坚长同学

Brief history of neural network



*张均瑜同学

第六次作业

题目 1

请判断以下关于 CNN 说法的正误，并解释原因：

1. 全连接层通常只出现在 CNN 的最后几层。

正确。全连接层通常作为分类器，通常只出现在 CNN 的最后输出部分。

2. 只有池化层能起到下采样的作用。

错误。下采样是指降低特征图或者输入图片分辨率的操作，除了

池化以外，步长大于 1 的卷积，插值等等也可以实现下采样。

3. 残差网络采用跳跃连接计算一个非线性函数使得输入能传递到更深的层。

错误。残差网络中的跳跃连接是一个恒等映射，不是非线性函数。

题目 2

假设输入是一张 128×128 的 RGB 彩色图像（3 通道）。

如果网络第一个卷积层的卷积核为 3×3 的大小、步长为 1、padding 为 1，共有 64 个卷积核，后接一个步长为 2 的 2×2 平均池化层，随后再接两层核大小为 5×5 、padding 为 2、步长为 1、卷积核个数为 128 的卷积层，以及一层步长为 2 的 2×2 的最大池化层，最后接含有 128 个神经元的全连接层和一个含有 10 个神经元的输出层。

试估算该网络的参数量，并回答网络的参数量主要由哪部分贡献；

（Optional）了解 CNN 的感受野及其计算。

参考答案：

1. 参数量计算

➤ 考虑偏置

$$\text{卷积层 1: } (3 \times 3 \times 3 + 1) \times 64 = 1792$$

$$\text{卷积层 2: } (5 \times 5 \times 64 + 1) \times 128 = 204,928$$

$$\text{卷积层 3: } (5 \times 5 \times 128 + 1) \times 128 = 409,728$$

$$\text{全连接层 1: } (32 \times 32 \times 128 + 1) \times 128 = 16,777,344$$

全连接层 2: $(128 + 1) \times 10 = 1290$

总参数量: $1792 + 204928 + 409728 + 16777344 + 1290 = 17,395,082$

➤ 不考虑偏置

卷积层 1: $3 \times 3 \times 3 \times 64 = 1728$

卷积层 2: $5 \times 5 \times 64 \times 128 = 204,800$

卷积层 3: $5 \times 5 \times 128 \times 128 = 409,600$

全连接层 1: $32 \times 32 \times 128 \times 128 = 16,777,216$

全连接层 2: $128 \times 10 = 1280$

总参数量: $1728 + 204800 + 409600 + 16777216 + 1280 = 17,394,624$

2. 参数量主要由哪部分贡献

全连接层

3. CNN 的感受野及其计算

自顶向下计算, $l_k = l_{k-1} + [(f_k - 1) * \prod_{i=1}^{k-1} S_i]$, 其中 l_k 表示第 k 层的感受野, l_{k-1} 表示第 $k-1$ 层的感受野, f_k 是第 k 层的卷积核大小, S_i 是第 i 层的步长。

感受野初始值 $l_0 = 1$, 这里只计算卷积层和池化层的感受野:

卷积层 1: $1 + (3 - 1) = 3$

池化层 1: $3 + (2 - 1) \times 1 = 4$

卷积层 2: $4 + (5 - 1) \times 1 \times 2 = 12$

卷积层 3: $12 + (5 - 1) \times 1 \times 2 \times 1 = 20$

池化层 2: $20 + (2 - 1) \times 1 \times 2 \times 1 \times 1 = 22$

题目 3

请在网上查阅以下网络，每种网络用两到三句话概括其主要思想以及核心技术贡献：

- DenseNet
- ResNeXt
- MobileNets/MobileNet V2/MobileNet V3
- EfficientNets

参考答案：

• DenseNet

主要思想：通过更密集的特征连接充分利用特征层，相比ResNet进一步减少参数，提高性能；由多个Dense Block组成，并且后面每个层都会接受其前面所有层作为其额外的输入，每层只学习很少特征图
核心技术贡献：节省参数和计算，抗过拟合性强

• ResNeXt

主要思想：借鉴Inception的“分割-变换-聚合”策略，但用相同的拓扑结构组建模块；每个模块内部同构多分枝减少超参数；引入“基数”，基数增加提升模型效果。
核心技术贡献：网络结构简明 模块化，手动调节超参小；准确度高

• MobileNets / MobileNet V2 / MobileNet V3

主要思想：把卷积核拆分并进行逐点卷积；v2加入了残差结构并设计了线性瓶颈（在出口、入口处用小通道，瓶子内部用大通道增大映射空间，将relu替换为线性函数；v3引入轻量级注意力模型，新的激活函数h-swish(x)及结合NAS与Net Adapt的网络结构搜索
核心技术贡献：基于流线性架构的深度卷积减少参数，使网络适用于移动端

· EfficientNets

主要思想：通过一起缩放 baseline 模型，同时调整深度、宽度、输入图片分辨率

核心技术贡献：对于任意有限的目标资源扩充最佳网络并保证模型有效性

*林少军同学

题目 4

1. 简述生成对抗的主要思想并简要分析 GAN 的目标函数。
2. 简述原始的 GAN 有哪些问题？
3. 简要梳理 GAN 发展至今的若干代表性工作，绘制发展路径图，每个工作用一句话总结其核心贡献（不限于本课程内的各种 GAN 变体）。

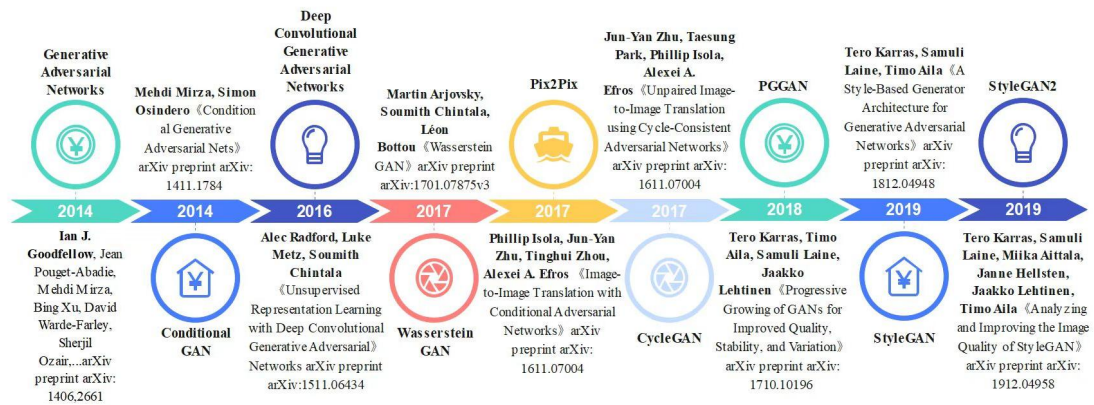
参考答案：

(1) 生成对抗的主要思想：生成器和鉴别器两个网络彼此博弈，生成器用于生成尽量逼真的假样本，鉴别器用于尽可能准确区分输入是真实样本还是生成的假样本。

GAN 的优化目标函数：
$$\min_G \max_D V(G, D) = \min_G \max_D E_{x \sim \text{data}} [\log D(x)] + E_{z \sim p_z} [\log (1 - D(G(z)))]$$

(2) 原始 GAN 的问题：

- ① 训练早期阶段，目标函数式无法为生成器提供足够大的梯度
- ② 最优鉴别器下，生成器的损失函数为常数，存在梯度消失问题
- ③ 采用技巧改进版本的生成器存在训练梯度不稳定，惩罚不平衡导致的模式坍塌，不好判断收敛性，难以评价生成数据的质量和多样性

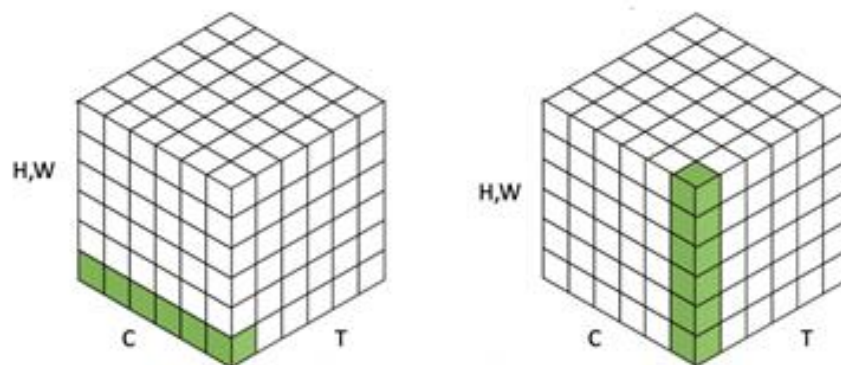


*张均瑜同学

第七次作业

题目 1

以下两个图表示的分别是什么类型的注意力机制？



参考答案：

左图：通道上的注意力机制；

右图：空间上的注意力机制。

题目 2

自注意力机制定义了哪三个关键元素？它们之间是什么关系、起

到什么作用？

参考答案：

三个关键因素：K，Q，V；key 和 query 通过点乘的方式获得相应的注意力权重，最后把得到的权重和 value 做点乘得到最终的输出。

题目 3

如果要采用 GCN 进行建模，针对化合物分类这项应用，其输出应该是 图 级别的；针对微博网站社交媒体账号分类这项应用，其输出应该是 节点 级别的。GCN 中主要靠 邻接/Laplace 矩阵的作用使得卷积能够方便的在不规则的图数据结构上运算。

题目 4

请问常规卷积和图卷积有哪些相同点和不同点？试各回答三点。

参考答案：

- 相同点：1) 局部连接；2) 权值共享；3) 多层网络。
- 不同点：

常规卷积	图卷积
定义在欧氏空间	定义在非欧氏空间
节点的邻域数量固定	节点的邻域不固定
卷积核大小固定	卷积核大小不固定

第八次作业

题目 1

与经典梯度下降法相比,进化算法更适用于哪些问题? 简述原因。

参考答案:

进化算法更适用于（1）多峰优化问题；（2）存在平坦区域的优化问题；（3）NP 难优化问题；（4）没有目标函数的优化问题。

梯度下降法是用一个点沿梯度下降方向进行搜索，本质上属于贪婪算法，全局搜索能力很差；进化算法同时用多个个体（即一个种群）协同对解空间进行搜索，具有较强的全局搜索能力和较高的搜索效率。所以进化算法不容易陷入局部最优（故适合问题 1），也能较好地搜索平坦区域（故适合问题 2）。也正因为此，对于 NP 难问题，进化算法能够在相对较短的时间，给出较满意的解，即接近最优解的解。

同时，进化算法是用 fitness 函数度量个体的好坏，这使得其即使在不存在目标函数的场景下，进化算法也能工作。

题目 2

用遗传算法实现背包问题，简述设计思路。

参考答案：

设有 N 种物品，第 i 个物品质量为 w_i ，价值为 v_i ，背包限重 W ，求怎样在不超重的情况下，背包中装的物品价值最高。

1. 基因编码：共有 N 种物品，每种物品只装一个或不装，分别用 1 和 0 代表。 N 件物品的选择就可以用 N 位二进制数表示，如 011001 ($N=6$)

2. 创建初始种群：随机生成 M 个个体的染色体，即随机生成一个由 1 和 0 组成的 $M \times N$ 矩阵

3. 适应度计算：设染色体 $x_1, x_2, x_3, \dots, x_{M-1}, x_N$ ，则适应度函数 $f(x)$ 为
$$f(x) = \sum_{i=1}^N x_i v_i, \text{ 且 } \sum_{i=1}^N x_i w_i \leq W$$
 适应度值 $eval(v_k) = f(x)$

4. 选择：采用轮盘赌法选择个体

① 计算适应度值；

② 计算所有染色体适应度之和： $F = \sum_{k=1}^M eval(v_k)$

③ 对染色体 v_k ，计算选择概率 $p_k = \frac{eval(v_k)}{F}$

④ 计算累积概率 $q_k = \sum_{i=1}^k p_i$

旋转轮盘 M 次，每次按下列步骤选择染色体构成新种群

① 生成 $[0, 1]$ 之间的随机数 r

② 若 $r < q_1$ ，选择 v_1 ；否则，选择 v_k ，满足 $q_{k-1} < r \leq q_k$

5. 交叉和变异：设交叉率 p_c 和变异率 p_m ，有 $p_c M$ 的染色体在断点处交换基因序列，有 $p_m M$ 的染色体在某基因位序列发生变化

6. 迭代：重复 3-5 操作，迭代 n 代，第 n 代中适应度最高的染色体序列即代表问题的解， $value_{max} = \max[eval(v_k)]$

*马艺同学

第一步，编码：基于背包问题的模型，我们将待求解的各量 $\{x_1, x_2, \dots, x_n\}$ 表示成长为 n 的二进制字符串 $x[j]$, $j=1, 2, \dots, n$ 。 $x[j]=0$ 表示物体 j 不放入背包内， $x[j]=1$ 表示物体 j 放入背包内。例如：1010代表一个解，它表示将第1、3号物体放入背包中，其它的物体则不放入。

第二步，生成初始种群：这里我们不妨将群体规模的大小取为4，即群体由4个个体组成，每个个体可通过随机的方法产生。例如：

$a_1 = [1, 1, 1, 1]$

$a_2 = [1, 0, 1, 0]$

$a_3 = [1, 0, 0, 1]$

$a_4 = [1, 0, 0, 0]$,

第三步，种群适应度计算：按照下列公式计算种群中个体适应度：

$$TotalSize = \sum s_i x_i$$

$$Fitness = \begin{cases} \sum v_i x_i & TotalSize \leq C \\ \sum v_i x_i - \alpha * (TotalSize - C) & TotalSize > C \end{cases}$$

公式的下半部分即为适应度的惩罚函数，其中参数 $\alpha > 1.0$ 。这里我们取 $\alpha=2$ 。

$a_1 = [1, 1, 1, 1]$, $totalSize = 14$, $totalValue = 19$, $fit = 9$

$a_2 = [1, 0, 1, 0]$, $totalSize = 6$, $totalValue = 8$, $fit = 8$

$a_3 = [1, 0, 0, 1]$, $totalSize = 7$, $totalValue = 10$, $fit = 10$

$a_4 = [1, 0, 0, 0]$, $totalSize = 2$, $totalValue = 3$, $fit = 3$

第四步，选择：我们采用与适应度成正比的概率来确定各个个体复制到下一代群

第五步，交叉：采用单点交叉的方法，其具体操作过程是：

1. 先对群体进行随机配对；
2. 其次随机设置交叉点位置；
3. 最后再相互交换配对染色体之间的部分基因。

在这里，

b_1 与 b_4 在第3位后交叉，生成：

$c_1 = [1, 0, 1, 1]$, $totalSize = 11$, $totalValue = 15$, $fitValue = 11$

$c_2 = [1, 1, 1, 0]$, $totalSize = 9$, $totalValue = 12$, $fitValue = 12$

b_2 与 b_3 在第2位后交叉，生成：

$c_3 = [1, 0, 1, 0]$, $totalSize = 6$, $totalValue = 8$, $fitValue = 8$

$c_4 = [1, 0, 0, 1]$, $totalSize = 7$, $totalValue = 10$, $fitValue = 10$

第六步，突变：我们采用基本位变异的方法来进行变异运算，其具体操作过程是：

1. 首先确定出各个个体的基因变异位置；
2. 然后依照某一概率将变异点的原有基因值取反。

在这里， c_1 的第1位发生突变：

$c_1 = [0, 0, 1, 1]$, $totalSize = 9$, $totalValue = 12$, $fitValue = 12$

$c_2 = [1, 1, 1, 0]$, $totalSize = 9$, $totalValue = 12$, $fitValue = 12$

$c_3 = [1, 0, 1, 0]$, $totalSize = 6$, $totalValue = 8$, $fitValue = 8$

$c_4 = [1, 0, 0, 1]$, $totalSize = 7$, $totalValue = 10$, $fitValue = 10$

至此，我们已经找到了2个最优解 c_1 与 c_2

*张晓怡同学

题目 3

你认为粒子群算法和蚁群算法中最关键的分别是什么？两者的异同点体现在哪些地方？

参考答案：

粒子群算法最关键的可以是：粒子的数量/更新的方式/加权的方式等等；蚁群算法最关键的可以是：信息素的设计/蚂蚁的数量等等。

- 相同点：都是群智能算法，模拟了生物群体的智能行为，都有较强的全局寻优能力；
- 不同点：一个模拟鸟群或蜂群，通过群体中个体之间的协作和信息共享来寻找最优解；一个模拟蚁群，在信息素的指引下在正反馈的作用下集中到最佳的路径上.....