

认知心理学基础

第七讲 记忆

主讲人：刘剑毅
人工智能学院

提纲

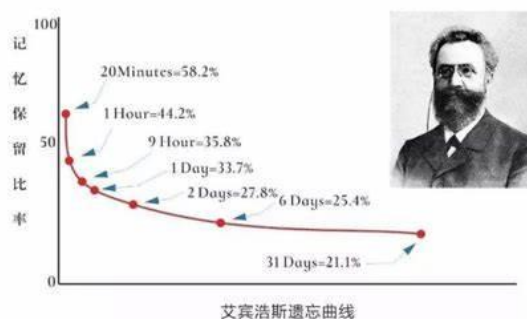
- 一. 记忆的结构模型
- 二. 短时记忆理论
- 三. 长时记忆理论

什么是记忆

- 人通过知觉从外界获得信息，再在记忆中贮存下来。
- 记忆是指人们编码、存储、提取过去的经验，并将这些信息用于当前情景的心理过程。
- 记忆将人的心理活动的过去、现在和未来联成一个整体，使心理发展、知识积累和个性形成得以实现。

记忆研究的历史

- 19世纪下半叶德国心理学家 Ebbinghaus 开创记忆研究，长时记忆的研究一直占主导地位
- 1890 年 William James “dual memory”
- 1965年 Waugh & Norman 两种记忆说和短时记忆的遗忘实验
- 1968年 Atkinson & Shiffrin 记忆的三级模型

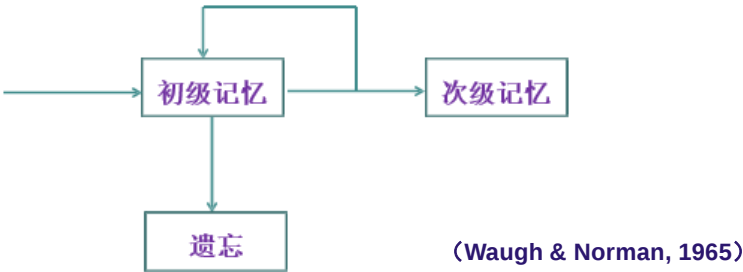


提纲

- 一. 记忆的结构模型
- 二. 短时记忆理论
- 三. 长时记忆理论

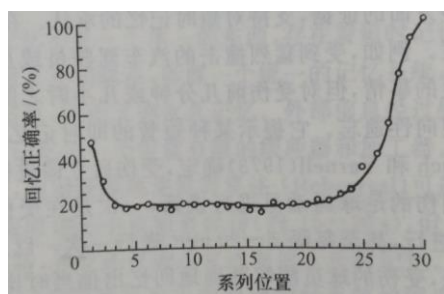
早期记忆存储结构模型

- 早期William James的双结构模型
- Primary memory (transitory) Vs. Secondary memory (permanent)

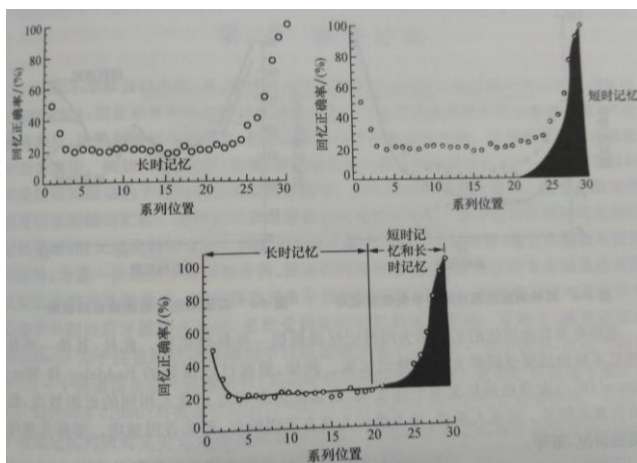


短时记忆存在的证据

- 自由回忆实验（Ebbinghaus）：被试尽量回忆刚刚呈现的刺激，但不必按照呈现时的顺序，原先的位置标识于横轴



IAIR Est. 1986



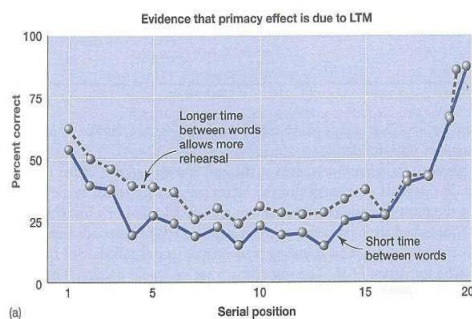
7

自由回忆实验

- Glanzer, Cunitz (1966)
- 实验范式设计

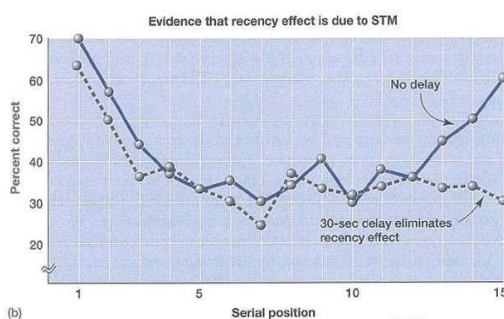
刺激呈现时间由1s增为2s

延缓回忆：刺激后心算作业30s再回忆



(a)

验证了长时记忆的存在



(b)

验证了短时记忆的存在

IAIR Est. 1986

2024/12/25

8

为两种记忆说提供了证据

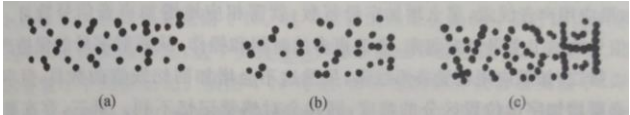
	为什么会发生？	它如何能被改变？
首因效应 序列位置曲线开头的 单词容易被记住	这些词语在展示单词列表 时被复述（ rehearsal ）所 以进入了LTM	为了增强首因效应，更慢地展 示单词列表，增加练习时间
近因效应 序列位置曲线结尾的 单词容易被记住	这些单词仍然存在于STM	为了减弱近因效应，在单词列 表消失后再等待30秒后，使 STM中的信息消失

记忆的多重结构

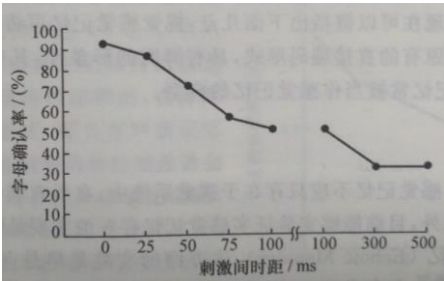
- 短时记忆：存储容量很有限，易受干扰，衰退快——秒为单位
- 长时记忆：存储容量相对无限而且保持时间也可以很长——时、日、年为
单位
- 感觉记忆：又称为感觉登记、瞬时记忆，每一种感觉贮存保持的时间均非
常短暂而且是感觉通道特异性的（即只局限于一种感觉通道）——ms为单
位
 - 视觉滞留现象：为什么电影的帧率是24Hz？

感觉记忆——视觉记忆

- Erikson,Collins(1967)的图像记忆 (Iconic Memory) 实验
- 实验范式设计:
先后呈现 (a) 与 (b) , 改变时间间距, 统计被试看出VOH的成功率

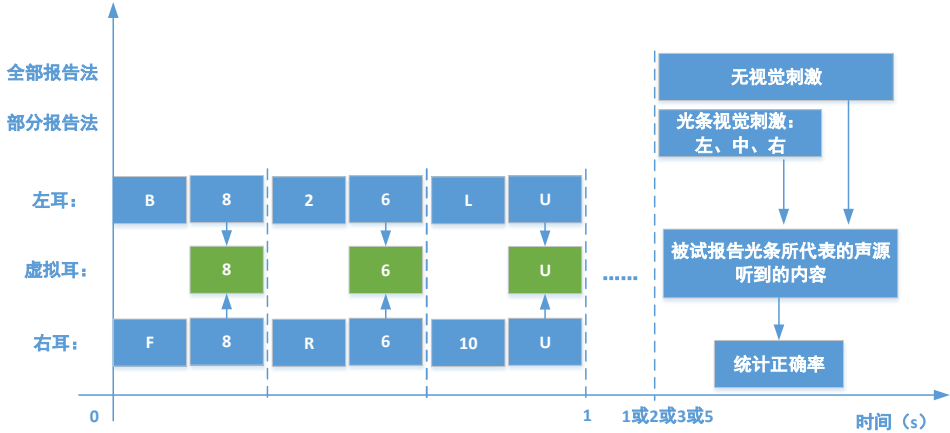


- 实验发现:
视觉感觉记忆的作用时间在300ms左右



感觉记忆——声象记忆

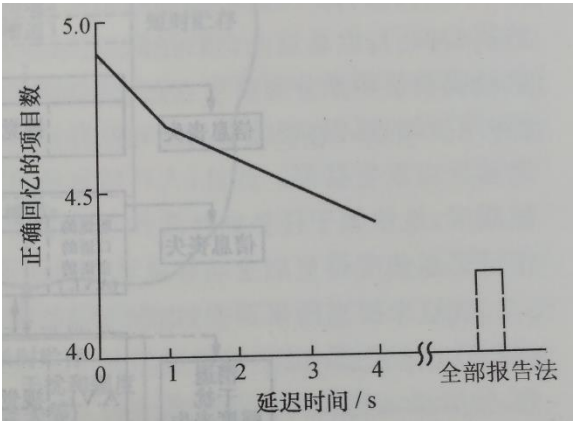
- Darwin (1972) 的 “三耳人” 声象记忆 (Echoic Memory) 实验
- 实验范式设计:



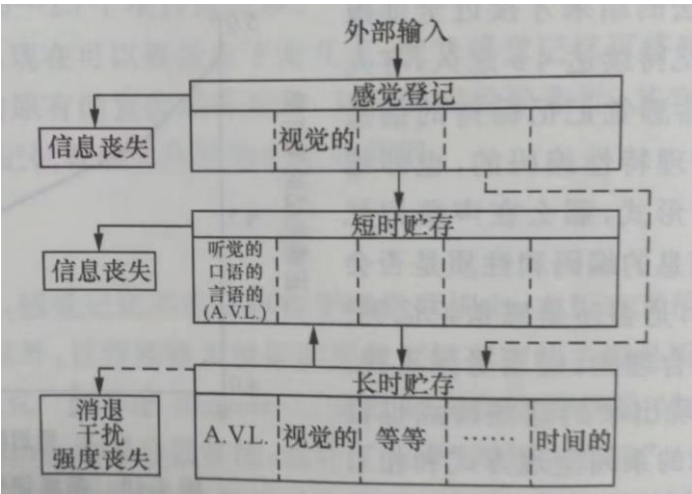
实验结果

实验结果:

- 部分报告法正确率均优于全部报告法
- 声象记忆容量为5个左右，而视觉记忆容量8~9个
- 声象感觉记忆的作用时间在4s左右，显著长于视觉记忆的几百毫秒

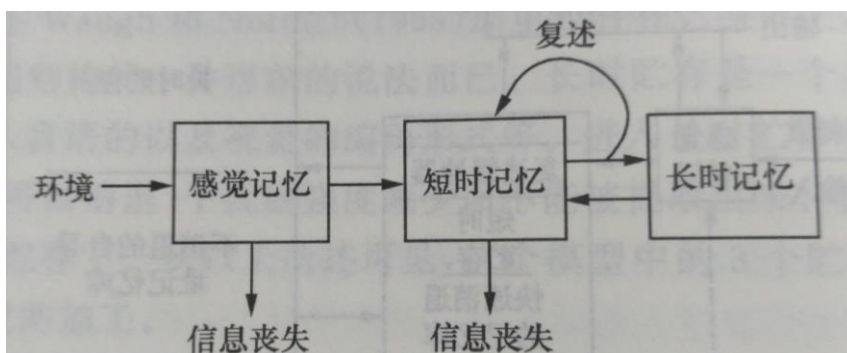


Atkinson & Shiffrin 三级记忆模型



简化的记忆信息三级加工模型

■ 目前认知心理学界最流行的记忆结构



记忆的加工水平理论

- Craik & Lockhart (1972) 对两种记忆说提出了尖锐的批评，提出了完全相反的假设：即“信息加工——记忆”，区别于之前的“记忆——信息加工”
- 认为记忆并不包括三种或几种彼此独立的存储系统，存储是在编码深度这个维度上连续变化的。信息被存储的水平取决于被加工的深入程度：
 - 浅层加工：只涉及很少的对含义的注意
 - 深层加工：涉及深层次的注意，专注于意义，并把它和其他事物关联

Craik 和 Tulving 的实验 (1975)

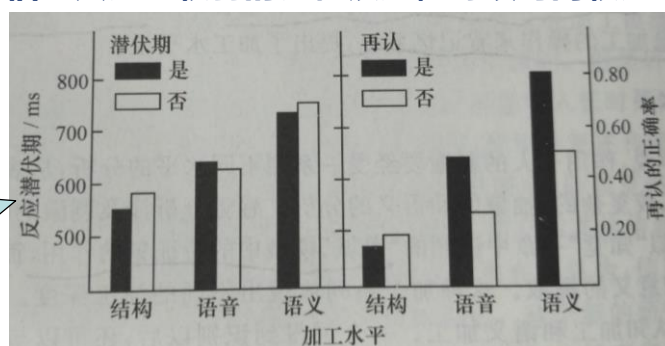
- 设计了一种不随意学习 (Incidental Learning) 的新的实验范式:
- 事先不要求记忆, 先行任务完成后突然通知被试进行回忆测试——“突击考试”
- 先行任务范例:
 1. Shallow processing: A question about physical features of the word
Question: Is the word printed in capital letters?
Word: bird
 2. Deeper processing: A question about rhyming
Question: Does the word rhyme with *train*?
Word: pain
 3. Deepest processing: A fill-in-the-blanks question
Question: Does the word fit into the sentence "He saw a _____ on the street"?
Word: car.

实验结果

实验结果及分析:

- 记忆痕迹是信息加工的副产品, 痕迹持久性是加工深度的函数
- 言语刺激比其他感觉刺激易得到更深的加工
- 只受到表浅分析的信息只产生较弱的记忆痕迹, 持续时间较短

被试做出回答所需时间体现了加工深度

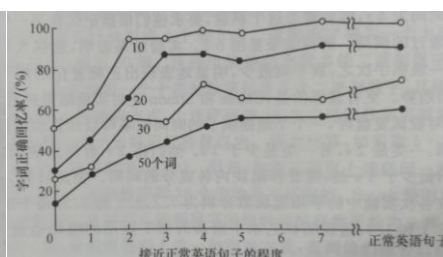
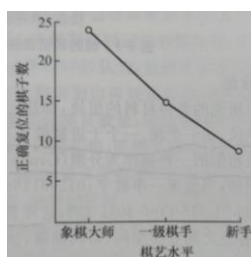


提纲

- 一. 记忆的结构模型
- 二. 短时记忆理论
- 三. 长时记忆理论

短时记忆

- 短时记忆容量：
- George A. Miller发表著名论文《神奇数 7 ± 2 ：我们信息加工的能力和限制》，指出人在一次阅读后，可以记住约7个无意义的字母、数字、音节、或字词
- 短时记忆的信息量不是以信息论中的bit为单位，而是以组块（Chunk）为单位
- 人利用贮存于长时记忆的知识，对进入短时记忆的信息进行合并，构成有意义的和熟悉的较大的单位
- 据Simon估计，象棋大师记忆的组块即棋局数约为5~10万个，积累的时间不少于10年



分组 (Grouping)

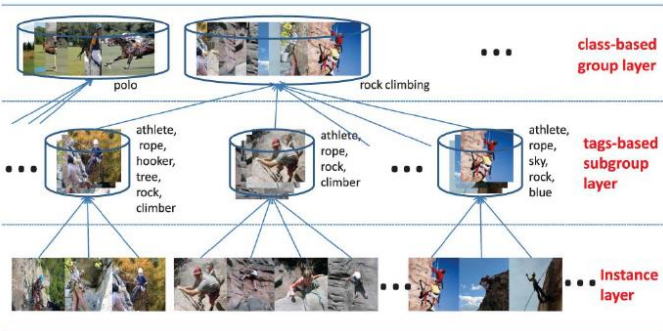
- 组块是有意义的，而即使无意义的组合，即分组，同样有利于短时记忆
- 例如：记忆电话号码 3528707537 vs 352 870 7537
- 分组是时间、空间接近项目的合并；**分组内部并不存在意义联系**，也不构成一个熟悉的单位
- 分组可以增加短时记忆容纳的项目，但作用远小于组块

基于组稀疏的图像标记 (Annotation)

- 组稀疏的基本思想是在完整码字L1范数约束的基础上，利用各类别子码的L2范数约束来使得码字尽可能趋近同类别表达
- 例子:在L1范数都是3的前提下，组稀疏编码的L2范数为 $\sqrt{3}$ ，小于普通稀疏编码时3的L2范数:

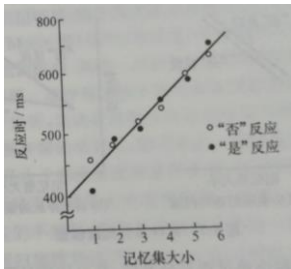
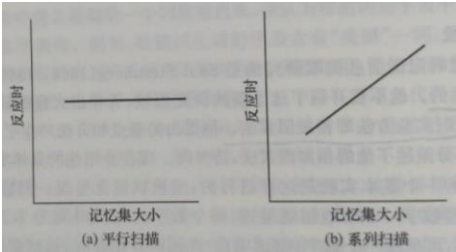
$$\sum_{i=1}^K (\|X_i\|_2 + \|Y_i\|_2) + \lambda (\|X\|_1 + \|Y\|_1)$$

类别1	类别2	类别3	
0 1 0 0 0	1 0 0 0 0	0 0 0 0 1	稀疏编码表达
1 1 0 0 1	0 0 0 0 0	0 0 0 0 0	组稀疏编码



短时记忆信息提取 (Retrieval)

- Sternberg的经典实验
- 实验范式：相加因素法。呈现一系列数字，然后被试判断一个新的数字是否刚才见过
- 实验变量：识记项目的数量
- 提出假设：平行加工或系列加工



- 实验结果：证实了短时记忆信息提取是系列加工的
- 序列加工的建模： $RT = e + cN + d$
 - 编码耗时
 - 记忆项目数
 - 决策耗时
- 线性回归结果： $RT = 38N + 397$

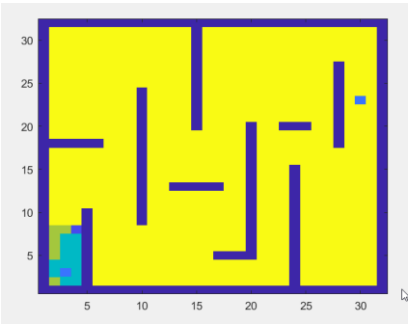
短时记忆信息提取 (Retrieval)

- 实验结果的讨论
 - ◆ 从头至尾扫描的假设：(for i=1:N 模式)
 - ◆ 自我停止扫描的假设：(while do 模式)
- 以上两种情况，作出“否”的反应需将识记项目 (Probe) 与记忆集 (Gallery) 中每个项目逐次比较
- 作出“是”的反应
 - ◆ (for i=1:N 模式)：与“否”的情形相同
 - ◆ (while do 模式)：最少比较1次，最多比较N次， $E(\text{比较次数}) = \frac{N + 1}{2}$
- 因比较次数影响线性回归的斜率，实验结果未出现1/2斜率现象，证实了从头至尾扫描假设

短时记忆信息提取 (Retrieval)

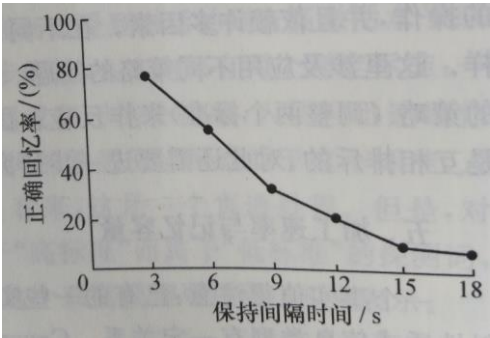
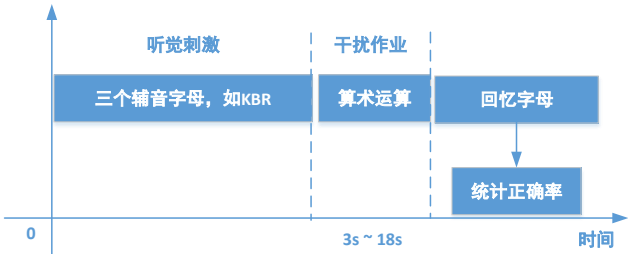
- 思考：为什么短时记忆要做这种似乎无效的工作？
- 从头至尾扫描：类比于**最优解**（如A*算法）
- 自我停止扫描：类比与**满意解**（如A算法）

■ 短时记忆的**从头至尾系列扫描**假设在字母、字词、颜色、面孔、听觉、不同被试身份的实验中均得到了验证，引起巨大反响



短时记忆中的遗忘

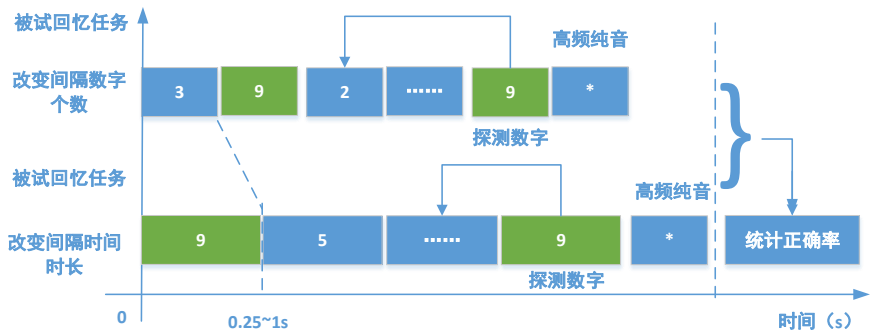
- 痕迹消除理论
- “Brown-Peterson”实验范式：在刺激呈现和回忆之间插入干扰作业，目的是为了阻止复述



- 实验结果
- 短时记忆保持时间短暂，如未得到复述，将迅速遗忘

短时记忆中的遗忘

- 干扰理论
- 痕迹消除 (fade) 和干扰 (interfere) 都可能是遗忘的原因，如何将这两个因素解耦却是困难的
- Waugh & Norman (1965) 设计了一个新的实验范式——探测法，巧妙实现了两个因素间的解纠缠 (disentangle)



IAIR Est. 1986

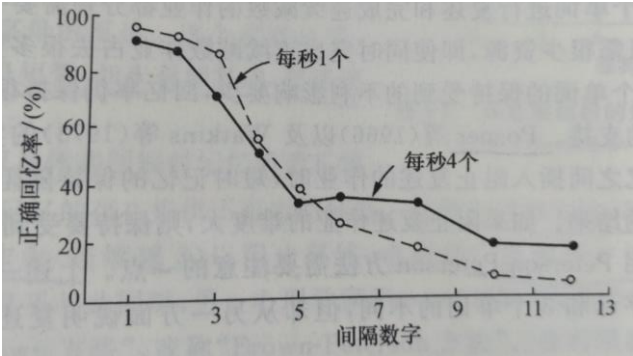
2024/12/25

27

短时记忆中的遗忘

- 解纠缠的原理
- 如果痕迹消除假说成立，则回忆正确率将随间隔时间的延长而减少
- 如果干扰假说成立，则回忆正确率将随间隔数字的增加而减少

- 实验结果
- 改变间隔时间未产生显著性差异
- 增加间隔数字个数在两种情况下均显著降低了正确回忆率
- 实验结论：短时记忆遗忘的主导因素是干扰而非痕迹消除



IAIR Est. 1986

2024/12/25

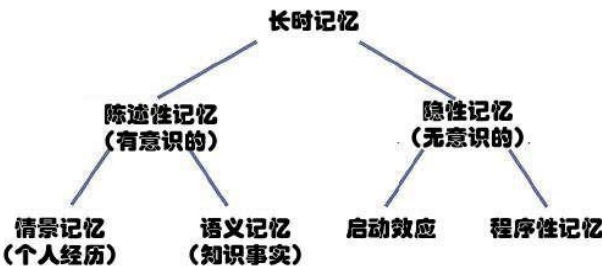
28

提纲

- 一. 记忆的结构模型
- 二. 短时记忆理论
- 三. 长时记忆理论
 - 3.1 长时记忆的分类
 - 3.2 长时记忆的模型

长时记忆

■ 长时记忆是真正的信息库，有着巨大的容量，存贮着我们关于世界的一切知识，使得我们能够识别模式，学习，运用语言、推理、解决问题..... 从而把人的活动的过去、现在和未来连成一个整体



You are what you know.

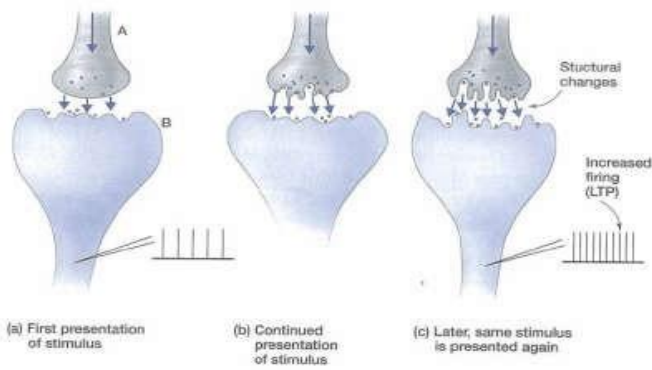
——Albert Einstein

长时记忆的神经生理学机制

- 突触
- 突触在信息存储中具有重要作用
- 赫布定律：描述了突触可塑性的基本原理，即突触前神经元向突触后神经元的持续重复的刺激可以导致突触传递效能的增加。这一理论由唐纳德·赫布(Donald Hebb)于1949年提出
- 突触的激活引起一系列的化学反应，导致突触产生新的蛋白质，并导致突触的结构改变

长时记忆的神经生理学机制

- 当一个刺激不断被重复时，突触产生结构改变，并导致激活频率的增加，这种激活频率的增加被称为“长时程增强效应”（long-term potentiation, LTP）



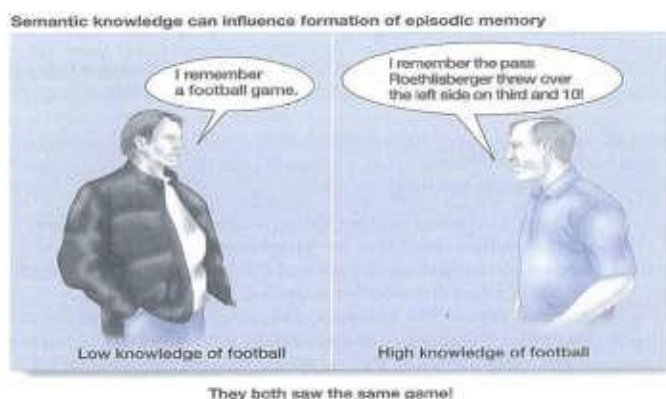
陈述性记忆

- 陈述性记忆是对有关事实和事件的记忆，它可以通过语言描述和传授而实现。
- 情景记忆是对个人在特定的时空背景下特定的经验或事件的记忆
- 语义记忆是个人关于世界的一般知识的记忆

区别方面	情景记忆	语义记忆
贮存内容	个人知觉经验	单词和概念的基本意义
类比	自传体	百科全书
参照框架	个人经历，时空	一般知识
稳定性	变化，易受干扰，提取难	稳定，不受干扰，提取易
推理性	推理能力小	推理能力大
提取	有意识的努力	自动化

语义记忆影响情景记忆

- 尽管两个人看得是同一场足球比赛，因为对于足球的知识的差异，他们会记住比赛中的不同的东西

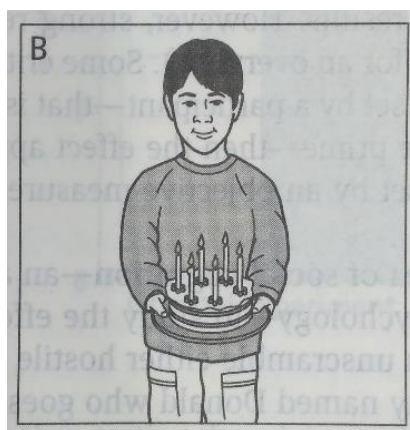
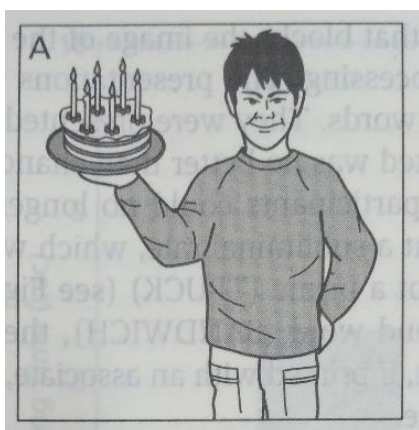


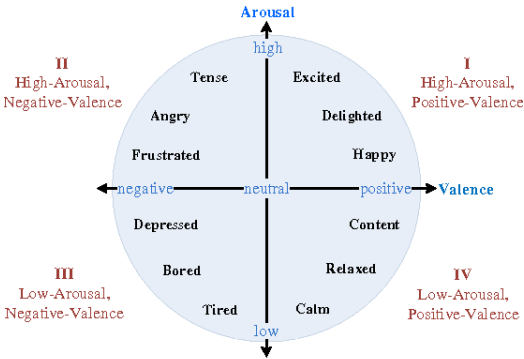
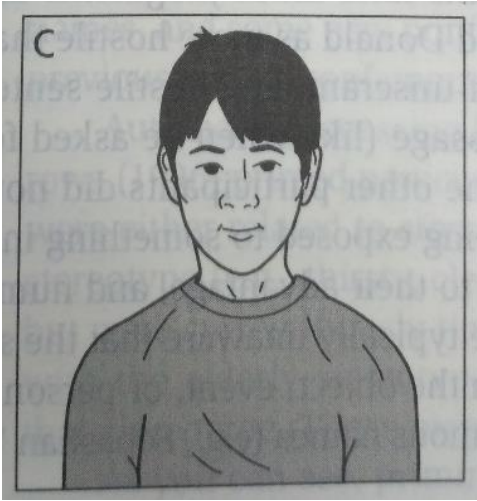
外显和内隐记忆

- 外显记忆是指在意识参与的条件下，过去经验对当前作业产生的有意识的影响。又叫**受意识控制的记忆**
- 内隐记忆是指在个体无法意识的情况下，过去经验对当前作业产生的无意识的影响，又叫**自动的无意识记忆**
- 测验方法
 - 外显记忆测验：要求被试有意识地回忆他在实验早期阶段学习过的材料的测验，比如自由回忆、线索回忆和再认
 - 内隐记忆测验：不要求被试有意识的去提取信息，只要求他们专注于眼前的作业，如知觉辨认，词干补全，残词补全等

启动效应 (Priming)

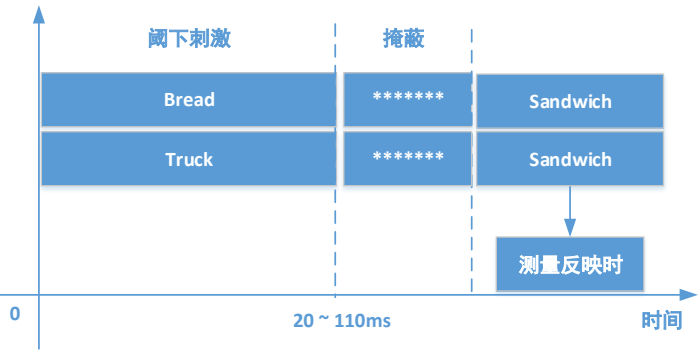
- 先前信息潜在影响了后续任务的表现，并且人可能不曾意识到回忆过先前信息，这就是**内隐记忆 (Implicit memory)**，与外显记忆相对应
- Meyer (1971) 的**词汇判断任务 (LDT)** 实验范式至今仍被采用。被试接受college这一视觉刺激后，相比接受jelly的刺激，对靶子词university的辨认速度更快
- 生活中经常会遇到“心理暗示”的情景
- 人几乎意识不到内隐记忆，它是启动效应研究的基础





Marcel (剑桥大学) 的阈下启动实验

- 采用了LDT实验范式设计
- 掩蔽的作用：破坏视网膜上的视觉暂留，避免短时记忆的影响
- 阈下刺激：呈现时间非常短，被试甚至未察觉到刺激
- 实验结果
- 关联词启动的反映时小于无关联词启动的反映时



程序性记忆

- 程序性记忆是针对“怎样做”或“如何做”事情的记忆，它以技巧性动作为基础，包括知觉技能、运动技能和认知技能的记忆
- 我们不会记得我们在何时何地学会了一些基础技能，但尽管如此，我们通常不会在使用这些技能的时候遇到麻烦
- 人们可以做事而不必意识到他们自己是如何做的，比如骑自行车
- 程序性记忆一旦形成，不会因为长时间不提取而被遗忘，比如游泳、开车

提纲

一. 记忆的结构模型

二. 短时记忆理论

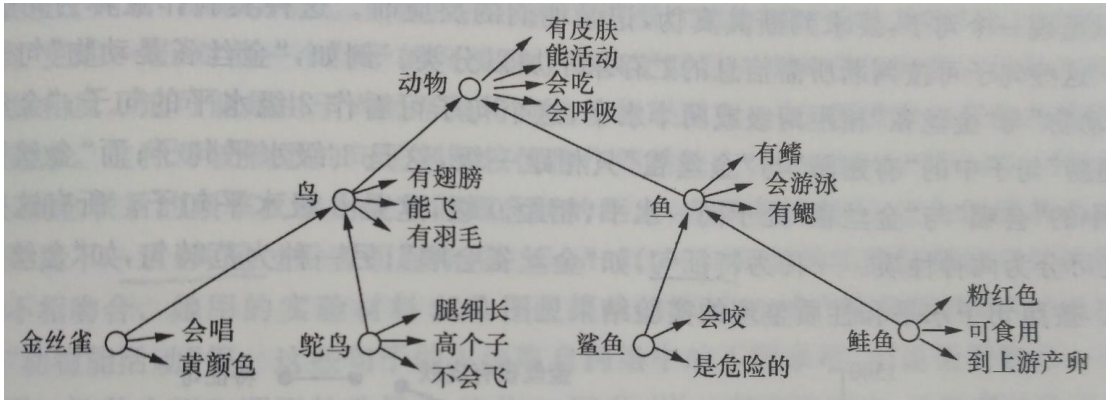
三. 长时记忆理论

3.1 长时记忆的分类

3.2 长时记忆的模型

长时记忆的模型

■ 层次网络模型 (Collins & Quillian, 1969)



IAIR Est. 1986

长时记忆的模型

层次网络模型的实验验证

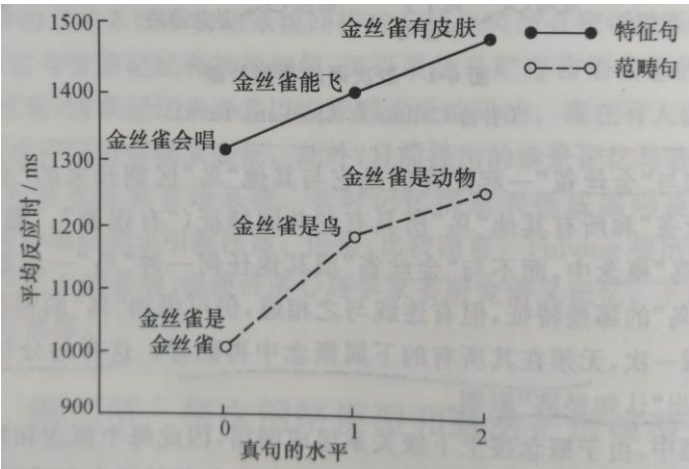
■ 实验范式设计：测量判断句子真伪的反应时

■ 实验发现：

相距0级判断需时最少，相距2级判断需时最长

■ 实验结果解释：

记忆搜索中，经过连线越多，需时越多



IAIR Est. 1986

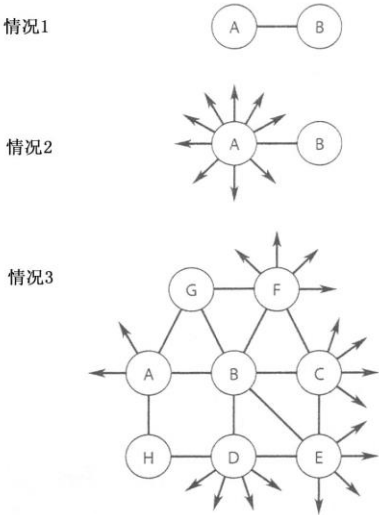
激活扩散模型

- 激活扩散模型 (Spreading Activation Model)是以语义联系或语义相似性将概念组织起来的网络模型
- 一个概念的意义或内涵由与它联系的概念来确定
- 概念特征不一定分级贮存，每个名称结点与语义网络中的多个结点相联系
- 当一个概念被加工或受到刺激，在该概念结点就激活，然后激活沿该结点的各个连线同时向四周扩散



扇面效应

- 由于从一个给定的结点发出的激活总量是一定的，如与该结点相联的结点越多，则激活分配越分散，每个结点所得到的激活越少，因此提取就越慢和越困难，这就是扇面效应——符合能量守恒原理
- 情况2中A启动B的难度大于情况1——扇面效应
- 情况3中B的激活最快，表明扇面效应可以通过增加知识间的联系来克服



激活扩散模型的验证——语义启动效应

- 回答下列问题：(Freedman & Loftus的经典实验,1971)
- 情况1：说出一种水果名称是以A字母开头
- 情况2：说出一种以A字母开头的水果名称
- 实验发现：情况1的反应时显著快于情况2；如再要求说出一种以P开头的水果名称（情况3），比情况1还要快
- 结果解释：水果这个概念与其下级概念“Apple”等有紧密联系，扩散到“Apple”等结点的激活既多又快，接近其活动阈限，所以一旦接收到“A字母开头”就可很快出现激活交叉，说出“Apple”；但是，在情况2中，由于扇面效应，反应要慢；情况3中，因为增加了知识间的联系，可以克服扇面效应，如说出“pineapple”

语义启动效应实验

- 词汇判断实验 (Meyer & Schvaneveldt, 1971)：呈现第一个字母串后，判断第二个字母串是否为英文单词

无 关	相 关
Nurse（护士）	Bread（面包）
Butter（黄油）	Butter（黄油）
940ms	855ms

- 激活（联想）扩散促进了词汇加工的速率
- 激活计算
- 读完三个词后联想：
- 圣经、动物、洪水
- 读完三个词更容易联想到诺亚方舟，而单独看一个词的联想则不同
- 语义记忆提取的速度和可能性由记忆的激活水平决定，而激活水平由基线水平激活和从相关概念获得的激活总和共同决定

特征模型

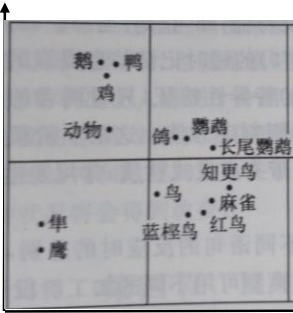
- 层次网络模型与激活扩散模型都属于网络模型范畴
- 特征模型范畴主要包括以下两种：
- 集理论模型（Set Theoretic Model）：概念由属性集或者特征集组成，**诸特征同等重要**，两个集之间的重叠程度越高，两个概念越接近——类比于汉明（Hamming）距离
- 特征比较模型：将两个集的特征进一步细化为定义性特征（mandatory）和特异性特征（optional），**且特征按重要性排列**，定义性特征高于特异性特征——类比于主成分分析（PCA）

语义空间

	知更鸟	鸟
定义性特征	是动物	是动物
	有羽毛	有羽毛
	红胸	—
	—	—
	—	—
特异性特征	会飞	会飞
	栖息树丛	—
	野生	—
	体小	—
	—	—
	—	—

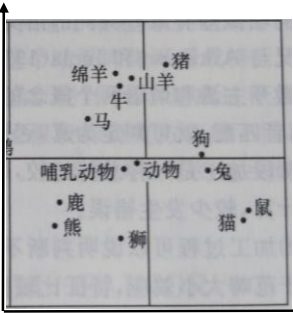
定义性特征与特异性特征

驯化程度



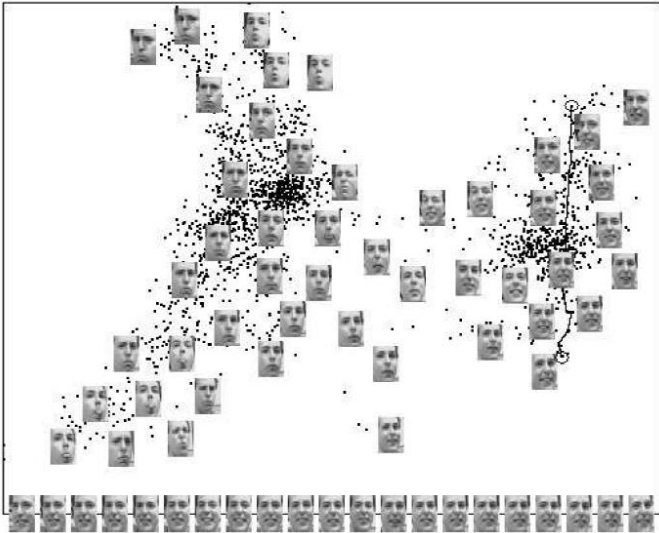
“鸟” 的二维语义空间

驯化程度



“哺乳动物” 的二维语义空间

流型 (Manifold) 学习



HAM模型

■ 人的联想记忆 (Human Associative Memory, HAM) 模型，也是一种网络模型，以命题作为知识的基本单元，表征复杂知识在长时记忆中的贮存和加工；通过4种联想适当结合，形成一个命题树，集语义记忆和情景记忆于一体，更具综合性。

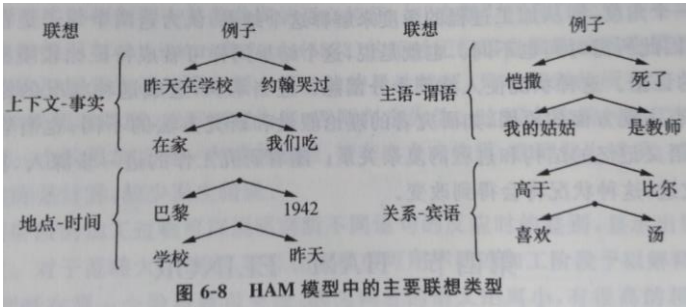


图 6-8 HAM 模型中的主要联想类型

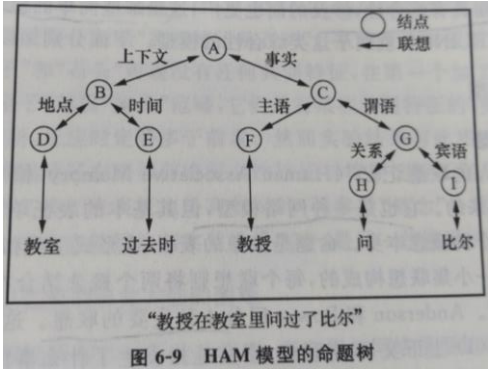


图 6-9 HAM 模型的命题树

记忆系统与无记忆系统

- 在任何时刻，系统的输出都只与当前时刻的输入有关，而与该时刻以外的输入无关，则称该系统是**无记忆系统**，否则就是**记忆系统**

例如：

$$y(t) = \frac{1}{C} \int_{-\infty}^t x(\tau) d\tau \quad (\text{电容}) \quad y(t) = x(t-1)$$

$$y(n) = \sum_{k=-\infty}^n x(k) \quad (\text{累加器}) \quad \text{RC、RLC电路}$$

$$y(n) = x(n) - x(n-1) \quad (\text{差分器}) \quad \text{等都是记忆系统}$$

无记忆系统的例子

- 设有一个系统 $y(n) = \sum_{k=-\infty}^{\infty} x(k)h(n-k)$ ，如果该系统是无记忆的，则在任何时刻 n ， $y(n)$ 都只能和 n 时刻的输入有关，上式中只能有 $k=n$ 时的一项为非零，因此必须有： $h(n-k)=0$ ， $k \neq n$ ，换元可得：

$$h(n) = 0, \quad n \neq 0$$

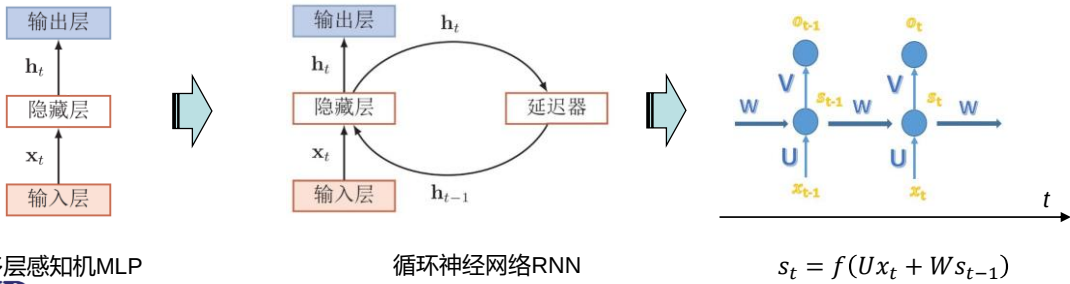
即该系统无记忆的充要条件为 $h(n)$ 是一个冲激函数：

$$h(n) = k\delta(n) \quad h(t) = k\delta(t)$$

不满足上述条件时，该系统是有记忆的。

循环神经网络RNN与短时记忆

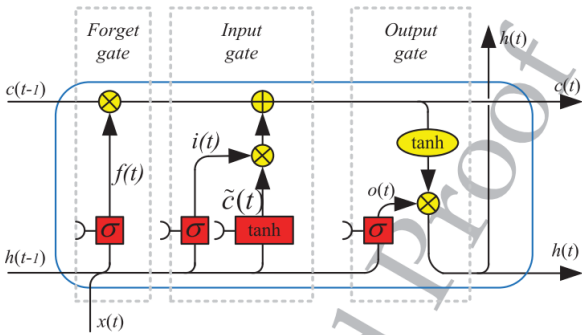
- 1933年，西班牙神经生物学家Rafael发现大脑皮层的解剖结构允许刺激在神经回路中循环传递，被认为是生物拥有短期记忆能力的原因
- CNN仅在层间建立权连接，而RNN在每层的神经元间建立连接，通过隐状态传递实现反馈效果，使得网络具有了记忆性，更适于处理时间序列数据，如语音识别、机器翻译等



IAIR Est. 1986

LSTM与长时记忆能力的数学实现

- LSTM是一种特殊的递归神经网络（RNN），设计用于解决传统RNN在处理长期依赖关系时遇到的梯度消失和梯度爆炸问题。LSTM通过引入门控机制来控制信息的流动，从而能够有效地保留和传递长期依赖关系。LSTM的核心结构包括输入门、遗忘门和输出门，这些门控制着信息的输入、遗忘和输出



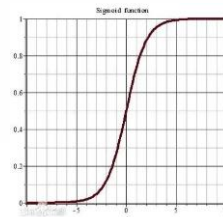
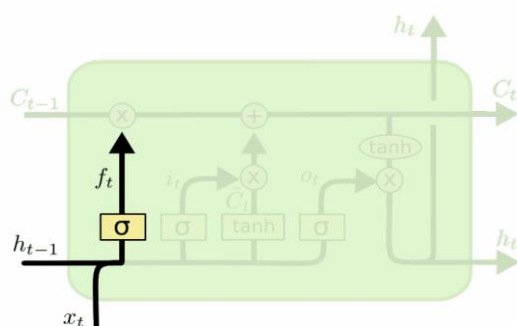
$$\begin{aligned} f_t &= \sigma(W_{fh}h_{t-1} + W_{fx}x_t + b_f), \\ i_t &= \sigma(W_{ih}h_{t-1} + W_{ix}x_t + b_i), \\ \tilde{c}_t &= \tanh(W_{ch}h_{t-1} + W_{cx}x_t + b_c), \\ c_t &= f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t, \\ o_t &= \sigma(W_{oh}h_{t-1} + W_{ox}x_t + b_o), \\ h_t &= o_t \cdot \tanh(c_t). \end{aligned}$$

IAIR Est. 1986

遗忘门 (Forget Gate)

- Sigmoid函数0~1间的输出值如果较小，能够通过乘法运算将记忆轴的数值进行抑制，实现遗忘功能

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$



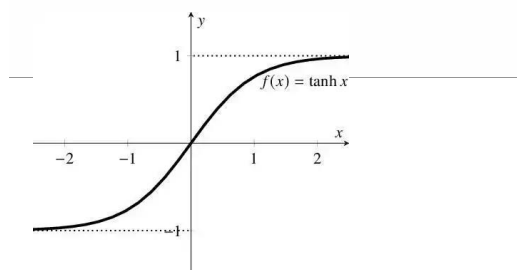
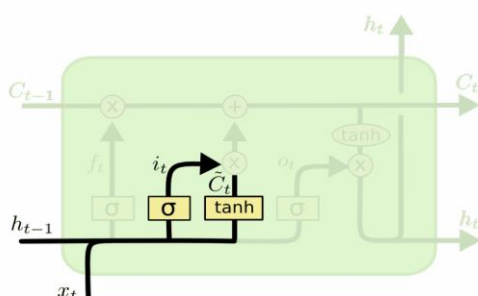
- Sigmoid函数的功能:
 - 将实轴映射为概率值 (0~1)
 - 避免输出值溢出
 - 可微函数实现求导的BP

输入门 (Input Gate)

- 输入门模拟接受外界刺激 (stimuli) 并通过加法被注入到记忆轴之中
- i_t 确保与遗忘门相同的数量级，实现可加性
- $\tilde{C}_t$ 不改变 x_t 的符号，确保输入的fidelity

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

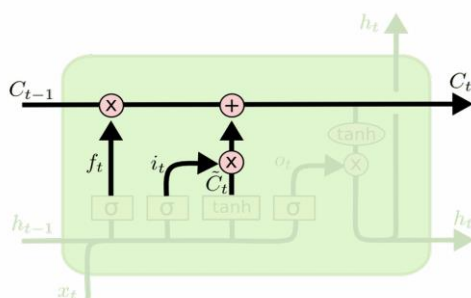
$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$



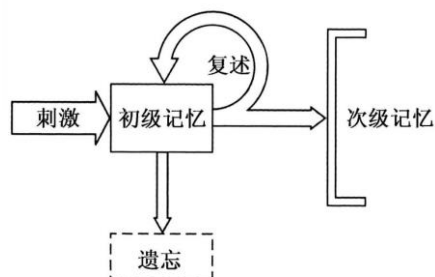
tanh解决了Sigmoid函数的不以0为中心输出的问题

记忆单元

■ 模拟短时记忆转为长时记忆



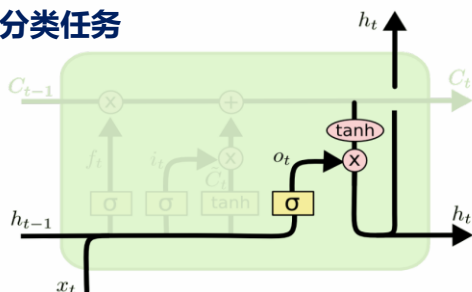
$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$



- 讨论：三个激活函数输出值大小与长时记忆、短时记忆、遗忘的关系？为什么重复的刺激（rehearsal）可以将短时记忆转为长时记忆？

输出门 (Output Gate)

- 综合外界刺激以及从长时记忆中提取（recall）的信息，对刺激做出响应
- tanh确保记忆提取信息的同号性
- 刺激对记忆提取信息进行加权，作为下一个隐状态
- 隐状态实现了对输入的特征提取（表征），同时作为输出，可以外接一个MLP实现分类任务



$$o_t = \sigma(W_o [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t * \tanh(C_t)$$

对比RNN的网络结构

- RNN是LSTM的一种简化形式，仅具备短时记忆能力；而LSTM在保留RNN隐状态传递机制的基础上，引入长时记忆轴线，从而能够有效地保留和传递长期依赖关系，并有效解决了梯度消失问题
- 下图3个cell其实是实例化后的一个对象（LSTM/GRU单元同理），以序列数据驱动的方式学习出cell中各个激活函数的一套参数，实现符合数据分布特点的长/短时记忆能力折衷

