

# 基于连通结构与动力学过程的知觉记忆层次模型\*

危辉<sup>1,2,3+</sup>, 栾尚敏<sup>4</sup>

<sup>1</sup>(复旦大学 计算机科学与工程系 认知算法模型实验室,上海 200433)

<sup>2</sup>(中国科学院 计算技术研究所 智能信息处理重点实验室,北京 100080)

<sup>3</sup>(浙江大学 计算机科学与工程学系 人工智能研究所,浙江 杭州 310027)

<sup>4</sup>(中国科学院 软件研究所,北京 100080)

## A Hierarchical Model for Perception Memory Based on Connected Graph and Dynamic Process

WEI Hui<sup>1,2,3+</sup>, LUAN Shang-Min<sup>4</sup>

<sup>1</sup>(Laboratory of Algorithm for Cognitive Model, Department of Computer Science and Engineering, Fudan University, Shanghai 200433, China)

<sup>2</sup>(Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, The Chinese Academy of Sciences, Beijing 100080, China)

<sup>3</sup>(Institute of Artificial Intelligence, Department of Computer Science and Engineering, Zhejiang University, Hangzhou 310027, China)

<sup>4</sup>(Institute of Software, The Chinese Academy of Sciences, Beijing 100080, China)

+ Corresponding author: Phn: +86-21-65110762, E-mail: weihui@fudan.edu.cn, <http://www.cs.fudan.edu.cn>

Received 2003-05-12; Accepted 2004-05-08

Wei H, Luan SM. A hierarchical model for perception memory based on connected graph and dynamic process. *Journal of Software*, 2004,15(11):1616~1628.

<http://www.jos.org.cn/1000-9825/15/1616.htm>

**Abstract:** One of the interferences between inheritance and concurrency is inheritance anomaly. From the view of cognitive computational neuroscience, a direct information representation method is presented based on neural system dynamics and graphic theory. A group of neurons and their connections representing perceptual information directly and the dynamical behaviors of neurons are defined firstly, and then a two-layer neural network is designed to record characteristics of stimulus and connect a specialized neural circuit that responding to the perception of that stimulus respectively. This could be achieved by the structure learning algorithm. The circuit constituted by neurons in two layers is also served as an associative memory of stimulus whose credibility is decided by the degree of

---

\* Supported by the National Science Foundation of China under Grant Nos.60303007, 60310213 (国家自然科学基金); the National Grand Fundamental Research 973 Program of China under Grant No.2002CB312103 (国家重点基础研究发展规划(973)); the Open Foundation of the Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, the Chinese Academy of Sciences, under Grant No.IIP2003-3 (中国科学院智能信息处理重点实验室开放基金)

作者简介: 危辉(1971—),男,江西南昌人,博士,副教授,主要研究领域为人工智能,认知科学;栾尚敏(1968—),男,博士,副研究员,主要研究领域为人工智能基础理论。

connection of the circuit. The direct representation method is of very significance to the research of semantic representation and inference driven by semantics in artificial intelligence.

**Key words:** associative memory; representation; computational neuroscience; dynamic system

**摘要:** 根据认知的计算神经科学的观点,提出了一种基于神经系统动力学理论和连通图的信息的直接表达方式.它首先定义了知觉信息直接表达的神经结构和动力学模式,然后提出一个双层的网络计算模型,分别用于记录外界刺激的特征信息和连通对应的特定神经回路的连接模式,这是通过结构学习来实现的.在两层神经元间建立起来的连通结构同时起到联想记忆的作用,记忆的可靠程度由神经回路的连通度来决定.这种直接表达方式对于人工智能中关于语义表达和基于语义的推理研究具有重要意义.

**关键词:** 联想记忆;表达;计算神经科学;动力学系统

**中图分类号:** TP18      **文献标识码:** A

记忆是整个心智的核心,它以感知为代表的低级脑功能活动和以言语为代表的高级脑功能活动具有基础性意义.由于脑结构与功能的极端复杂性,对记忆物质基础的研究是极其困难的<sup>[1,2]</sup>.目前,神经生物学界与生理心理学界对记忆机制的研究或是集中在细胞和分子水平上<sup>[3,4]</sup>,如突触传递效能、可塑性的生化基础等,或是集中在对参与记忆的脑区功能定位上<sup>[5,6]</sup>,如海马在学习与记忆过程中的作用等.它们分别处于微观和宏观两端,由于解剖学、电生理学、脑功能成像等实验手段上的不足,神经元输入与输出的信号投射路径的精细连接跟踪困难,所以处于中间层次的细胞网络组织上的记忆机制研究相对不足.而正是这个中间层次的研究<sup>[7,8]</sup>对计算机科学家或人工智能学家有巨大的吸引力,使得他们试图给出种种记忆算法模型来弥补这个差距.这种努力在人工神经网络方面的实践是极其丰富的<sup>[9]</sup>.

联想记忆的计算模型是人工神经网络研究的中心内容之一,时下该领域研究的焦点集中在联想记忆的模式<sup>[10,11]</sup>、局部极小问题<sup>[12]</sup>、权值与阈值的修改算法<sup>[13]</sup>、容量问题<sup>[14]</sup>等方面.然而在记忆的生理基础与心理过程研究、记忆的人工神经网络模型研究间存在的差距,导致许多记忆现象不能由给出的模型获得解释,比如重要的心理学发现:近期定律(law of recency)和重复定律(law of repetition).如果对比人的记忆实践与记忆的人工计算模型,我们会发现几个重要的问题:(1) 人的记忆系统对样本集合性不敏感;(2) 人的记忆系统对样本顺序性不敏感;(3) 人的记忆系统对样本模式一致性不敏感;(4) 人的记忆系统对样本的组合性不敏感;(5) 人的记忆系统对损伤仅局部性敏感;(6) 人的记忆系统对单一特殊样本的增加、删除、部分修正没有全局性影响;(7) 人的记忆系统在异联想上有感觉通道的模式性.这些事实虽然是现象上的,并且还无法看到脑内部导致这些特征出现的细节是什么,但有理由相信在以特征编码、直接表示、结构学习、激励扩散为基础的面向结构形态的记忆模型比面向权值调整的记忆模型更接近真实情况,对结构学习的研究<sup>[15-17]</sup>需要将微观的亚结构与宏观的集群联系起来,即将存在于分子与细胞水平上的信息加工机制与存在于脑功能分区水平上的信息加工机制联系成一个整体,科学意义明显.对于知觉信息表达的方法研究请参阅文献<sup>[18-20]</sup>.将记忆与表示问题结合在一起进行的研究正日益受到关注<sup>[21]</sup>.

本文所描述的就是基于已知的生物神经元信息传递与连接特性,面向知觉信息表达与知觉信息加工过程的记忆模型,它提出和具体完成一种可以对知觉信息进行直接表达且同时实现联想记忆功能的方法.第一,提出了一种基于神经系统动力学理论和连通图的信息的直接表达方式.它首先定义了知觉信息直接表达的神经结构和动力学模式,然后提出一个双层的网络计算模型,分别用于记录外界刺激的特征信息和连通对应的特定神经回路的连接模式,这是通过结构学习来实现的.在两层神经元间建立起来的连通结构同时起到联想记忆的作用,记忆的可靠程度由神经回路的连通度来决定.这种直接表达方式对于人工智能中关于语义表达和基于语义的推理研究具有重要意义.第二,实现了上述方法.实现过程是基于一个神经系统动力学和矩阵特征值和特征向量的数学模型.它首先设计了单个神经元模型的动力学特征及其通过将输入输出配对特征向量化的连通构建操作和对应的矩阵计算过程,然后给出对于整个网络的结构学习算法.它克服了基于权值改变的学习算法的一些缺陷.连通结构的建立在进行表达的同时也使得联想记忆得以实现.第三,还进行了回忆与纠错实验,以及作

为记忆载体的连通结构的规模与学习次数的关系,发现基于结构表达和学习的连通结构的建立速度较之连通结构本身规模的变化表现出稳定性,而且所需的学习步数相对较少,然后,对网络模型的容量和建立不同连通程度的连通结构的概率进行分析,发现能够以很高的概率保证在一定的学习步数内实现单侧弱连通,同时在增加学习步数和连通结构的规模的情况下也能以稳定的概率得到强连通结构.此模型可以解释许多记忆现象.

## 1 基于生物神经元信息传递及其组织特性的记忆层次模型设计

### 1.1 动力学过程的连通结构形式

神经元是人脑中唯一进行决策的结构,它们数量庞大,彼此间连接众多,生物物理学发现发生在它们内部和之间的复杂的电学过程完成了信号的变换与传递,这表明神经系统是一个动力学系统.虽然脑神经系统动力学目前尚无公认的定义,但它的研究对象可定为从神经元到功能柱及其联结的时空结构和演化机能,动力学过程的研究有助于了解脑的整体性原则<sup>[22-24]</sup>.动力学过程同时发生在不同的层次上,这取决于我们所观察的认知行为的精细尺度,就对某一外界刺激(如景物)产生知觉、建立记忆、形成回忆这样层次的问题而言,我们设想它们的动力学过程发生在一个局部的神经回路上,就是分布在各个层次间纵向局部和同层次内横向局部中的一个神经连接结构,这个连接结构表达和记忆特定的刺激,回路的形成、全部或部分活跃等微观的物理事件就对应到产生知觉、形成记忆或回忆等心理事件上.下面要提出的就是这样一个局部连接结构的形态、形成和动力学过程.

网络结构上的每一个节点都有自己的行为特征,它们接受、处理和发送信息,各自以特定的方式影响着与其相连的其他节点的行为,从网络的运行模式上来看,就是有许多信息在节点和它们的连接线上有向地流动.因此,用有向图来刻画这种网络结构及其运行模式是非常自然的.那么可以设想:对一个刺激物的知觉表达是一个特定的神经回路模式,对这个刺激物形成了记忆就对应于这个神经回路能建立稳定的连接和确定性地被再次唤起,而这个神经回路可以对应到一个有向图上,神经元的轴突和树突可视作连接线,如图 1 所示.

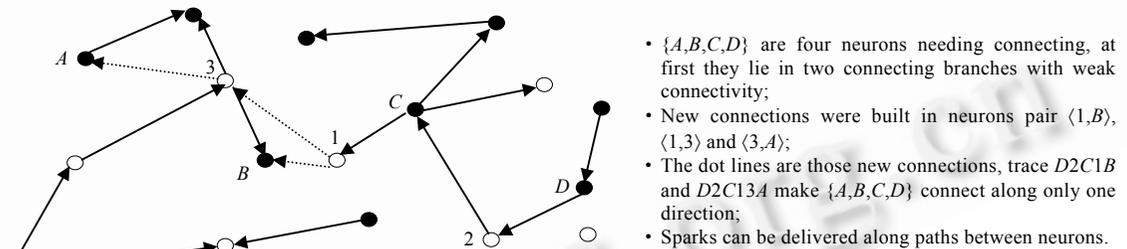


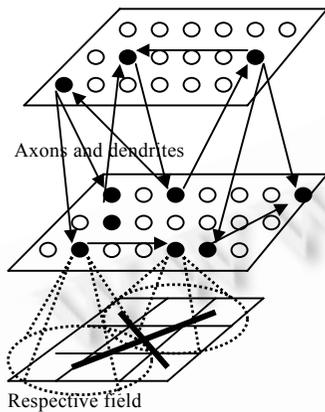
Fig.1 Formation of a connecting structure on a directed graph

图 1 有向图上的连通结构建立

### 1.2 随机连接的双层网络结构

在人的中枢神经系统中,神经元在功能上有分工,在分布上有层次性,在不同层间和同一层内存在着广泛的联系,这种等级组构对信息进行逐级的加工.图 2 就是这样一个受神经生理学启发的网络结构模型,主要讨论其上面两层,即特征响应神经元阵列和连通神经元阵列.此网络结构的设计有生理学依据,但由于是计算机的算法模型,所以图 2 中的结构并不专门对应于某一特指的生理神经组织.在图的 I 层和 II 层中设计了两种功能不同的神经元,分别起到沟通和记录刺激特征的作用;激励关系沿着已经存在的连接传播,新的传播途径由学习算法来添加;连通结构就在分属 I 层和 II 层的神经元间建立起来.即使位于同一区域的真实神经元所联系的其他神经元在数目和范围上也仅有大致的规律,所以基本上可以视为随机连接.此模型也是如此,因此和传统的人工神经网络结构(如 BP 网络)有很大的不同.此处,每个神经元都有不同的个性,需要个别处理,这也正是说它们表现出动力学特征的原因之一.生理神经元的树突和轴突随着学习和经验是经常发生变化的,表现为数量的变化或者形态的变化,这反映出它们本质上是通过对结构的改变来实现学习的.本文中设计的网络模型也是结构学习,

而不是权值学习,也可以说是用结构来表示,而不是用权值来表示.被记忆的东西对应于特定的结构,而不是对应于特殊的权值.传统的人工神经网络在结构和学习算法上存在着一些缺陷:(1) 神经元间的连接是固定的,学习算法所改变的是连接的权值,而不是连接的方向;(2) 学习样本带有顺序性,同一样本集合不同的学习排列顺序对最终学习结果有影响;(3) 学习样本间存在相关性,学习通常是对一个样本集合进行的,更换其中的一个或几个样本对最终的学习效果有影响,而这些相关性并不是语义性质的,或者说这种相关性的存在并不是依据逻辑的必然性;(4) 改变网络局部的连接或权值会产生整体性影响;(5) 学习算法的收敛速度慢.目前人工神经网络与神经科学的联系还较弱,网络结构僵硬、不灵活,权值的调整技巧很难预料与掌握.而本文中的网络模型就是针对这些缺陷的改进,也即结构学习的特点所在.当一个刺激物出现后,它激活了大量特征响应神经元阵列上的神经元,这些神经元又连锁激活连通神经元阵列中的其他神经元,这时需要这个联想记忆的网络模型判别:这个神经活跃模式是新的模式呢,还是旧的模式?前者需要记忆,后者实现了回忆.



- (a) Linking neurons array, they act as linker and distribute their axons and dendrites randomly in (I) and (II), which take part in the construction of connected graphs or neural circuits. All neurons and their connections are shareable, that makes association spreading over scale variable circuits possible. Axon outputs signals and dendrites bring out signals in.
- (b) Feature corresponding neurons array, it is similar to the hypercolumn in visual cortex or frequency neuron in aural cortex, which can be activated by special aspects of outer stimulus. Receptive field of these neurons falls in next layer, the topological shape and size of it is variable because neurons connected randomly.
- (c) Receptors array in sensing channel, they can be visual, aural or haptic sensors group, or their mixture. There is a stimulus drawn in it for the sake of demonstration.

Fig.2 A hierarchical model for associative memory

图 2 联想记忆的层次模型

### 1.3 单个神经元的动力学过程算法

在下文中,神经元专指在联想记忆双层网络模型中设计的人工计算单元,而不是指真实的生物神经元.各个神经元具有相似的形态特征和行为规律,但它们的参数各不相同.本节单个神经元的计算模型如图 3 所示.

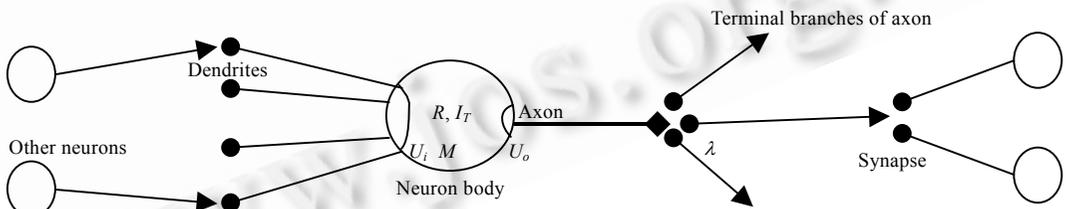


Fig.3 Structure characteristics of a neuron

图 3 神经元的结构特征

每个神经元有两种状态:休眠或活跃.一定强度的输入信号所具备的能量可以激活神经元,即从休眠状态变成活跃状态.当状态为活跃时,神经细胞体能进行一些操作.每个神经元有数量不定的树突以接收不同的输入信号,以及一个轴突(多个轴突末梢)来输出信号.图 4 做了简化,在不影响问题性质的前提下,假设所有的突触连接都在树突上;假设所有的神经元都受一个共同的离散时间脉冲的同步,信号的接收和发送由时钟使能(enable),整个网络的动力学演进步骤就由此共同时钟来协调.每个神经元活跃时都具有一定的能量  $E$ ,表示其活跃的程度,但在活跃期内会持续单调衰减,直至衰减到 0 而恢复成休眠状态.也可以说,神经元的活性随时间延长而递减,每个神经元都有一个用时钟周期个数来衡量的最长活跃期限. $U_i$ 表示信号输入端的电势, $U_o$ 表示信号输出端的电势,树突和轴突上是否有信号通过由它们两端的电势差决定,前提是方向不能反转(树突只承载输入,轴突

只承载输出).当神经元休眠时, $U_i$ 与 $U_o$ 是相等的;而当神经元由休眠变成活跃时,细胞体的活动会使电势 $U_i$ 降低,同时使电势 $U_o$ 升高, $U_i$ 与 $U_o$ 间保持 $-POT\_D$ 的电势差(potential difference),这样做的目的是使信号此刻更容易从输入端进入,从输出端流出;神经元具有一个简约电阻 $R$ ;任意两个通过轴突和树突在突触处传递信号的神经元间,其连接线上的电流这样来计算: $I_k=(U_o-U_i)/R$ ,其中 $U_o$ 和 $U_i$ 分别来自信号前后端两个神经元,且 $U_o \geq U_i$ .每个神经元都有要被激活所需的门限电流 $I_T$ ,若所有输入信号的总电流数 $\sum I_k$ 大于 $I_T$ ,神经元即由休眠状态进入活跃状态,否则维持休眠状态.连通矩阵 $M_{in}$ 和 $M_{out}$ 记录信号的流向,是结构学习的结果.它的维数是可变的,受树突数和轴突末梢数影响.神经元细胞体对信息的处理围绕这两个连通矩阵进行.图4演示了神经元在不同状态下各项属性的变化,其中的数学曲线只反映其变化的趋势,具体解析形式由各个神经元的特性定义.图中的中间段为活跃期,前、后两段为休眠期.

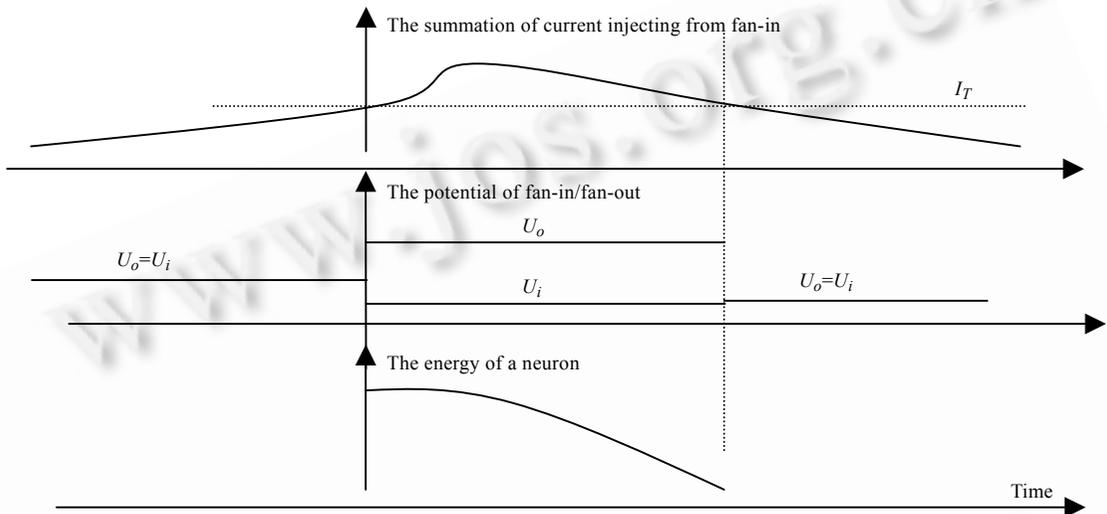


Fig.4 Variation of several items of a neuron under different status

图4 不同状态下神经元中各项属性的变化

单个神经元执行算法1,即单个神经元的动力学特征算法.

#### 算法1.

- (1) 在时刻0,神经元初始化各项参数,处于休眠状态,开始等待任意一个树突分支的信号输入.
- (2) 时刻 $t_1$ 有一个或数个树突分支有信号到来,若它们的强度之和小于等于 $I_T$ ,则继续等待,循环执行(2);反之,当现在输入的强度之和大于 $I_T$ 时,该神经元被激活,转向(3).
- (3) 记录被激活时刻为 $t$ ,神经元获得一个活跃的能量 $E$ ,降低信号输入端电势 $U_i$ ,对输入信号进行处理,升高信号输出端电势 $U_o$ ,将处理结果从轴突输出.
- (4) 检查 $E$ 是否耗尽,若没有,则维持输入端和输出端的电势,继续接受、处理和发出信号, $t=t+1$ ,持续活跃状态,循环执行(4);否则转向(5).
- (5) 当 $E$ 耗尽时,记录时刻 $t_2$ ,转入休眠状态,重新初始化该神经元各项参数,转向(2).

因为每个神经元都是高度可重用的,可参与对多个知觉状态的记忆,同一个神经元在不同的神经回路或在同一神经回路的不同时刻都扮演不同的角色:它的输入和输出来自不同的方向,即需要沟通不同的方向.所以,每个神经元都要记住不同时刻激励的传导走向.神经元的工作分为两类:学习与记忆、识别与回忆.这两类工作分别更改连通矩阵和使用连通矩阵.学习与记忆过程发生在先,当一个刺激物出现并激活大量的特征和连通神经元时,首先要在它们之间建立起一个连通结构.假设这个连通结构已经建立,如图5所示,画出了连通图的局部(虚线以上)以及其中一个神经元与其他神经元的连接关系.

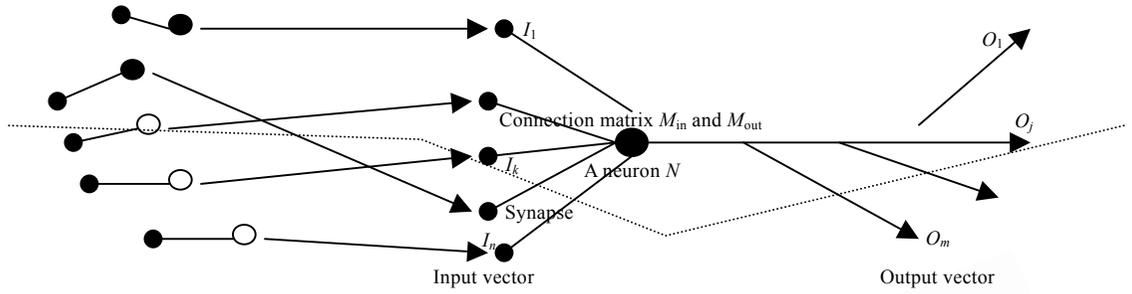


Fig.5 Function of a neuron in a connection structure

图 5 连通结构中一个神经元的作用

落实到每个活跃的神经元上的任务是记住:当且仅当如此这些树突有这样的输入时,哪几个轴突末梢有那样的输出.图 5 中神经元  $N, n$  是其树突的数目,  $m$  是其轴突末梢分支的个数,其所有输入端的输入组成一个输入向量  $V_{in} = \{I_1, \dots, I_k, \dots, I_n\}$ , 所有输出端的输出组成一个输出向量  $V_{out} = \{O_1, \dots, O_j, \dots, O_m\}$ , 其中  $I_k$  和  $O_j$  的计算如前所述, 如果  $I_k$  或  $O_j$  所代表的连接不属于正在被记忆的连通结构, 那么  $I_k$  或  $O_j$  的值就是 0. 在神经元的轴突上信号的强度是  $O = \sum_{j=1}^m O_j$ . 当一个尚处于学习或说被记忆阶段的、针对样本  $s$  的连通结构逐渐成型时, 任意一个神经元  $N$  的树突和轴突的输入/输出的分布在相对时刻  $t$  呈现为一个向量对偶, 称为路由对偶:  $route_N^{s,t} = \{V_{in}^{s,t}, O^{s,t}, V_{out}^{s,t}\}$ . 这样, 该神经元在其参与记忆的任意样本、任意时刻下所有的路由对偶必然组成一个按出现时刻先后的有序队列:  $route\_queue_N = \{route_N^{s,t} | \forall s \forall t\}$ . 实际上, 队列  $route\_queue_N$  不可能无限长, 我们认定在队列长度(设为  $L$ )全满的情况下后来的路由对偶会替代队列中被激活次数最少和相对最为陈旧的路由对偶, 换句话说, 被替代的也就被遗忘了. 神经元的任务是记住各自的队列  $route\_queue_N$ . 此时可以发现, 对任何一个神经元的路由对偶队列而言都是独特的, 任何一个对偶的 3 个分量也是一一对应的, 因此将集合  $\{\{V_{in}^{s,t}, O^{s,t}\} | \forall s \forall t\}$  和集合  $\{\{V_{out}^{s,t}, O^{s,t}\} | \forall s \forall t\}$  分别看作是由两个矩阵的特征向量与特征值组成的对偶的集合. 这两个矩阵就是连通矩阵  $M_{in}^N$  和  $M_{out}^N$ , 使它们满足:

(1) 由集合  $\{\{V_{in}^{s,t}, O^{s,t}\} | \forall s \forall t\}$  得到满秩矩阵  $IN = (V_{in}^{s1,t1}, \dots, V_{in}^{sL,tL})$ , 且  $IN^{-1} \cdot M_{in}^N \cdot IN = \begin{pmatrix} O^{s1,t1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & O^{sL,tL} \end{pmatrix}$ , 可

知  $M_{in}^N = IN \cdot \begin{pmatrix} O^{s1,t1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & O^{sL,tL} \end{pmatrix} \cdot IN^{-1}$ ;

(2) 由集合  $\{\{V_{out}^{s,t}, O^{s,t}\} | \forall s \forall t\}$  得到满秩矩阵  $OUT = (V_{out}^{s1,t1}, \dots, V_{out}^{sL,tL})$ , 且  $OUT^{-1} \cdot M_{out}^N \cdot OUT = \begin{pmatrix} O^{s1,t1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & O^{sL,tL} \end{pmatrix}$ , 可知  $M_{out}^N = OUT \cdot \begin{pmatrix} O^{s1,t1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & O^{sL,tL} \end{pmatrix} \cdot OUT^{-1}$ .

要使上述想法得以实现, 还必须确定以下 4 个细节. 第一, 鉴于矩阵计算的要求不妨设定每个神经元树突的数目等于其轴突末梢分支的数目, 记为  $n$ ; 两个连通矩阵  $M_{in}^N$  和  $M_{out}^N$  的阶数因此也为  $n$ ; 该神经元路由对偶队列  $route\_queue_N$  的长度  $L$  也等于  $n$ . 第二, 矩阵  $IN$  和  $OUT$  是由队列  $route\_queue_N$  累积起来的, 而且还要满秩, 即要分别存在  $L$  个线性无关的特征向量, 而理论上神经元未发生学习之前, 队列  $route\_queue_N$  应当是空的. 但一个神经元在形成之初总是有许多预制连接的, 这些连接的输入/输出排列数是巨大的, 当学习发生后队列  $route\_queue_N$  开始增长, 可以从预制的排列中随机选择若干对向量作为队列  $route\_queue_N$  剩余空间的补充, 使它们线性无关满足矩阵  $IN$  和  $OUT$  满秩的条件, 简单的做法是选择一些单位向量作为补充, 这是容易实现的. 若  $IN = (V_{in}^{s1,t1}, \dots, V_{in}^{sL,tL})$  满秩, 则又记为  $IN = (V_{in}^1, \dots, V_{in}^L)$ ; 同理, 若  $OUT = (V_{out}^{s1,t1}, \dots, V_{out}^{sL,tL})$  满秩, 则可记为

$OUT=(V_{out}^1, \dots, V_{out}^L)$ .随着学习的继续,学习得到的向量将逐步填充并替代队列  $route\_queue_N$  中随机选择的向量.第三,若新学习得到的向量与以前学习得到的向量线性相关,即输入向量  $V_{in}^i = \sum_{i=1}^L k_i V_{in}^i$ ,其中  $\{k_i\}$  是不全为 0 的常数集,就淘汰新向量,即放弃通过这个神经元建立连通的企图,还可以从其他神经元建立连通.大量连通神经元与神经连接的存在和结构可变学习保证了这一点的实现.但在实际使用中,若网络规模不大或  $n$  的数值较小,则发生线性相关的可能性和频率会增大,这时可以逐次增加向量的位数.这些位是虚拟的,  $V_{in}^i$  的虚拟位设为 1,其他  $V_{in}^i$  的对应虚拟位设为 0.这就消除了线性相关性,随之  $M_{in}^N$  的阶数也虚拟扩大.当然,只有向量的前  $L$  位发挥实际的控制流向作用.第四,由于连通矩阵  $M_{in}^N$  和  $M_{out}^N$  是构造出来的,所以通过简约电阻  $R$  的调整,可以避免矩阵的特征多项式方程出现重根的情况,使得特征矩阵  $O^{s,j}E - M_{in}^N$  和  $O^{s,j}E - M_{out}^N$  的秩为  $L-1$ .

学习与记忆过程是由多个特征向量和特征值求两个矩阵,反过来的回忆和联想过程是由一个矩阵的一个特征向量求另一个矩阵的另一个特征向量.由  $M_{in}^N \cdot V_{in}^i = O^i \cdot V_{in}^i$  求出  $O^i$ ,再由  $M_{out}^N \cdot V_{out}^i = O^i \cdot V_{out}^i$  求出  $V_{out}^i$ ,其中  $V_{out}^i$  表示的正是往后的激励传播路径.至此可以得到算法 2,即基于矩阵特征值的单个神经元结构学习算法.

#### 算法 2.

- (1) 如果神经元处于休眠状态,则等待,循环执行(1);否则,转向(2).
- (2) 神经元处于被激活状态,提取输入向量  $V_{in}$ ,如果  $V_{in}$  是  $M_{in}^N$  的特征向量,转向(3);否则转向(5).
- (3) 根据  $V_{in}$ ,  $M_{in}^N$  和  $M_{out}^N$ ,依前述方法求出输出向量  $V_{out}$ .
- (4) 回忆过程:执行单个神经元动力学特征算法(算法 1),维持  $V_{in}$  和  $V_{out}$  的连接关系,直到算法 1 结束或新的输入向量到来,均转向(1).
- (5) 记忆过程:记录此刻输出端要求建立连通的扇出请求,记为  $V_{out}$ ,将新得到的路由对偶  $route_N = \{V_{in}, O, V_{out}\}$  加入对偶队列  $route\_queue_N$ ,然后判断由新队列得到的新矩阵  $IN$  和  $OUT$  是否满秩;否则,依前述方法调整其中的向量对,使它们满秩,再求取新的连通矩阵  $M_{in}^N$  和  $M_{out}^N$ .
- (6) 执行单个神经元动力学特征算法(算法 1),维持  $V_{in}$  和  $V_{out}$  的连接关系,直到算法 1 结束或新的输入向量到来,均转向(1).

基于突触活跃组合排列的统计,将排列靠前的默认设定为特征向量,这是动态变化的,也即有概率因素参与.处理向量是矩阵优势,这也是选择矩阵运算作为神经元细胞体运算方式的原因.

### 1.4 自主的、并行分布式连通求取算法

在新生儿的脑生理发育过程中有一个突触的爆发生长期(bloom bursting),会产生极为巨量的神经连接.在新生儿以后的认知发展过程中,随着对外界刺激的加工的增加,这些神经连接会进行自行调整和裁剪<sup>[25]</sup>.不仅如此,认知发展成熟以后,神经元树突的形态也是经常变化的,使得神经元能够改变接收信息的对象的范围.这些显然都是对结构的学习.

在前一节中讲述的是关于单个神经元的动力学行为和结构学习行为,那么,作为知觉信息的直接神经表达模式,即整体上众多神经元的连通结构是怎样建立起来的呢?这就是自主的、并行分布式连通求取算法.它的核心就是建立连通结构,逐步发展到强连通结构.要知道,仅靠各个神经元独自反复运行算法 1 和算法 2,虽然可能实现局部的连通,但不能保证最后可以得到一个连通结构.一个未经过学习的网络中的连接是神经元间初始的预制连接,连通结构开始只能依靠这些,通过学习,要使连通程度得到提高.如果开始不能得到一个包含所有活跃神经元的连通结构,那么它们必定分布在几个连通分支中,通过寻找沟通神经元进行接力沟通.

所有神经元的活跃状态只能持续一段时间,如图 5 所示.连通神经元需要反复激励才能改变其连通特性,这有两个途径:第一,神经元的激活门限值以一定的概率随信号由输入端输入进行激活尝试次数的增加和输入强度的增加而降低;第二,根据发育神经科学,神经元输入野与输出野的范围发生变化,也即树突和轴突的形态会发生变化,从而导向需要连通的方向,这需要一个实际的物理场的帮助,在算法上可以用随机改变连接的办法来实现.现在详细说明第 1 种方法,即算法 3.1.

**算法 3.1.**

- (1) 计算输入端的输入强度,  $I = \sum_{k=1}^n I_k > 0$ , 如果  $I$  大于  $I_T$ , 则执行算法 1, 结束算法 3.1; 否则转(2).
- (2) 计算概率  $p = e^{-\beta I}$ ,  $\beta$  是一个常数, 产生一个随机数  $0 < a < 1$ , 如果  $p < a$ , 结束算法 3.1; 否则转(3).
- (3) 上凸的双曲函数的极限  $\lim_{x \rightarrow +\infty} \frac{e^x - e^{-x}}{e^x + e^{-x}} = 1$ , 以它为系数, 改变该神经元的激活门限值  $I_T = I_T - \frac{e^I - e^{-I}}{e^I + e^{-I}} \cdot I_T$ .

第 2 种方法很明了, 即算法 3.2.

**算法 3.2.**

(1) 如果一个沟通神经元有输入, 但它的输出端平静, 则此刻随机地改变(包括增加)该神经元的 1 个或几个轴突末梢; 结束.

(2) 如果它的输出端的其他神经元活跃, 而输入端平静, 则随机地改变它的 1 个或几个树突; 结束.

以上两种存在诱导的情况, 通过结构的改变使此神经元可能同时与两个连通分支建立连接. 算法 3.1 和算法 3.2 结合起来就得到诱导沟通神经元改变连通特性的算法.

最后, 经过前面几种算法的准备, 得到自主的、并行分布式连通求取算法, 即算法 4.

**算法 4.**

(1) 被激活的所有神经元各自进行激励的传导计算, 即算法 2, 若干步后, 若所有的活跃神经元都在一个连通结构中, 则转向(5); 否则, 可以得到几个连通分支.

(2) 位于各个连通分支内的神经元, 按照算法 3.1 改变其他神经元的门限值, 使连通分支外的神经元可能被激活而建立新的沟通, 实现连通分支的合并, 如果所有连通分支都合并了, 则转(5), 否则进行下一步.

(3) 得到若干经过合并后的连通分支, 按照算法 3.2 改变某些神经元的连通方向, 可能实现新的合并, 如果所有分支都合并了, 则转(5), 否则进行下一步.

(4) 若刺激存在, 维持激励, 再转(1); 否则, 算法结束.

(5) 连通求精算法: 反复执行算法 3.1 和算法 3.2, 建立可能的新连接, 由单侧连通或弱连通向强连通的方向发展, 直到刺激消失, 激励源消失为止, 算法结束.

算法 3.1、算法 3.2 和算法 4 不一定每次都能成功, 但只要刺激反复出现, 强化激励源, 使神经元活跃, 经过多次尝试就能使连通状况得到改善, 这人的记忆特征相符. 连通结构是否成功建立, 连通的程度如何, 要靠能否成功回忆来检验. 当给出原先刺激的局部时, 且其激活神经元位于整个弱连通结构的下位, 那么整个结构就不能被激活, 对知觉的神经表达模式回忆失败; 反之, 若是强连通结构, 则激活结构的任何局部都能唤醒整体, 回忆成功.

结构学习的优点是: 结构可变, 网络规模可在不影响已有学习成果的情况下扩充, 学习过程中任何一步都有明确的指向意义, 样本之间不会发生非语义因素的互相干扰, 且不存在学习的次序性, 学习速度快且学习效果可随时独立改进, 连通结构具有冗余度, 表达稳定牢靠, 任何局部损伤不影响其余部分的运行和其余记忆的结果, 生理和心理意义明显.

**2 实验结果及其分析**

在程序实现中, 设定特征神经元阵列的规模为  $10 \times 10$ , 沟通神经元阵列的规模为  $30 \times 30$ , 所有神经元树突和轴突末梢分支的数目均为 20, 同一阵列内和不同阵列间的树突与轴突的连接是在层次网络模型构建之初随机形成的. 所有神经元的内部信息和结构特征数据都保留在由 Access 的数据文件中, 通过 ODBC 进行访问. 图 6 是此联想记忆模型的程序界面, 其中有特征神经元阵列和沟通神经元阵列(整齐排列纯粹是为了方便观察), 以及对一个刺激进行的连通结构表达, 所建立的连通图和部分激励传播路径也显示了出来. 由于黑白打印的关系, 用于区别神经元特性的颜色已被忽略了.

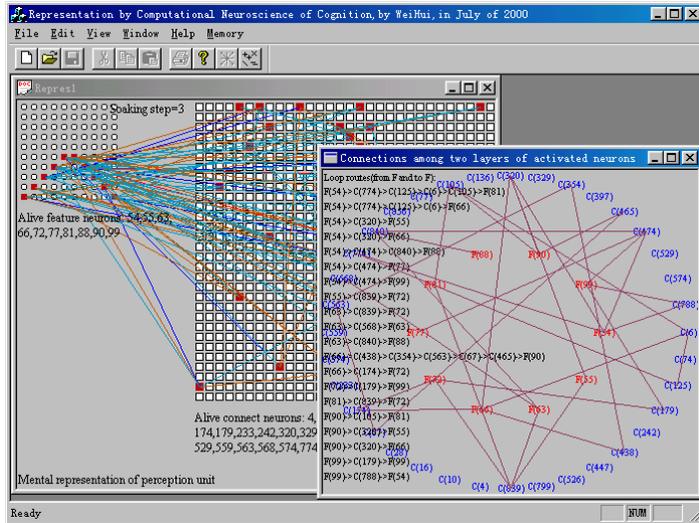


Fig.6 Interface of the associative memory model

图 6 联想记忆模型的界面

2.1 回忆与纠错实验

图 7 给出了一次联想记忆的实验结果.为了方便观察,假设被记忆模式由呈直线排列的 10 个特征神经元组成(如图 7(a)所示),它们经过 6 步结构学习形成一个连通结构(省略);图 7(b)为测试用的模式,假设它只有 4 个特征神经元活跃,但只有两个是属于原模式的有效激励,另外两个是多余的干扰;经过联想记忆,针对初始模式形成的连通结构被活跃的两个神经元激活了结构中的其他部分,并排除了干扰,成功地回忆出原来的模式(如图 7(c)所示),其中 0 表示休眠的神经元,1 表示被激活的神经元,与图 7(a)完全一样.

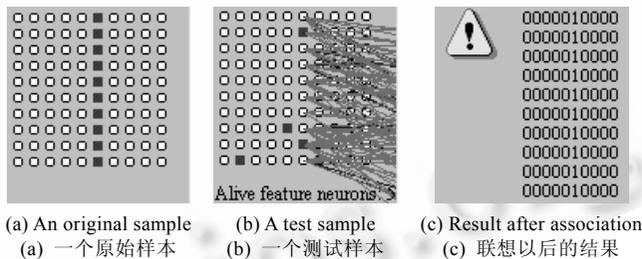


Fig.7 An experimental result of associative memory

图 7 联想记忆的一次实验结果

2.2 连通结构建立的速度与规模

图 8 是不同规模的连通结构学习次数、路径条数等指标的对比,其记忆样本为 20 个,其中分别用 10,15,20,25 个特征神经元来表达的样本各占 1/4.可见,学习次数稳定在 9 以下,基本上不随连通结构规模的扩大而变化,这是结构学习的优越性所在,其他指标基本上呈单调上升的趋势.这是由于联想学习的网络模型中预制连接的随机性均匀分布造成的.

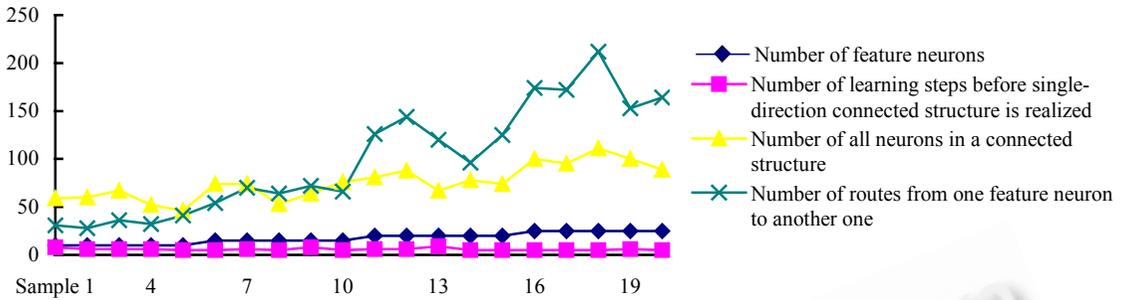


Fig.8 Number of learning steps of different connected structures

图 8 不同连通结构的学习次数

### 2.3 学习次数与连通程度的对比

图 9 显示了对应于图 7 所用连通结构的建立过程与学习次数之间的关系.左边是只学习了一次的情况,可见,所激活的特征神经元(由 F 标识,内圈)和连通神经元(由 C 标识,外圈)数目相对于右边要少,而且还没有出现由特征神经元到特征神经元的单侧连通路径;右边是经过 6 次学习以后的情况,激活的神经元数量有很大的增加,而且出现了许多路径.模式中所有 10 个特征神经元均出现在路径的末端,表示它们都可能被同一模式中其他活跃的特征神经元激活.

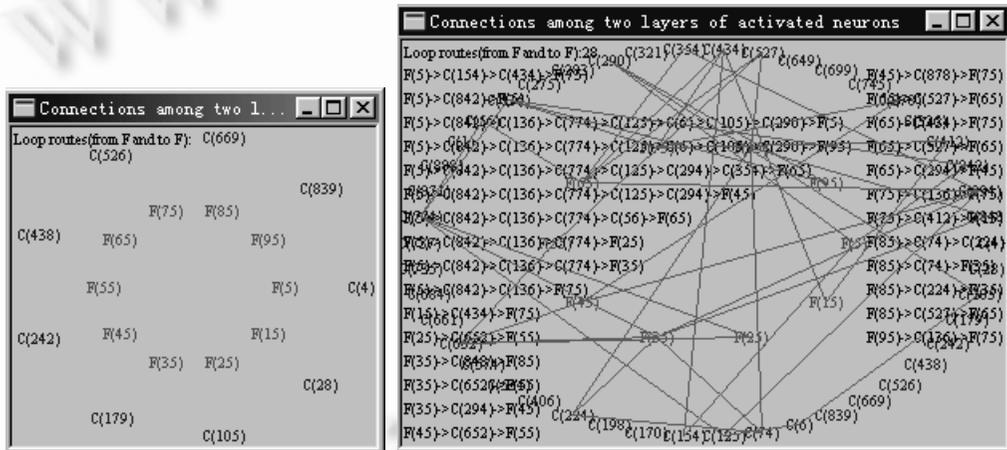


Fig.9 Figuration of connected structure

图 9 连通结构的建立

## 3 容量、纠错率分析及记忆现象解释

### 3.1 容量分析

由于此联想记忆网络模型中有动力学过程、随机连接、连通结构表示和结构学习等概念,所以其中可变因素非常多,要进行一个网络可以存储多少个样本模式这样的容量分析非常困难.结构表示和学习使得记忆模式所需要的神经元个数不同、连通形式不同,神经元还可以高度重用,因此只能在限制的情况下来讨论这个问题.

首先,假设已经设定好了预制连接,且不考虑第 4 节中采用的连通矩阵增加虚拟位的情况,那么每个连通矩阵  $M_{in}^N$  和  $M_{out}^N$  线性无关的向量数最多为  $n$ (沿用前文的设定),这样,每个神经元能稳定实现的沟通模式数最大就为  $n$ .然后,假设联想记忆的神经网络总数为  $M$ ,那么可能作为连通结构中一个节点的扇入/扇出形态的沟通模式总数为  $n \cdot M$ ,区分不同的神经表达模式就由不同的神经元和不同的扇入/扇出形态决定.由于神经元和彼此间

的连接都是可重用的,假设由  $k(0 < k < M)$  个神经元构成的连通结构,单从神经元的不同来看,可能的容量是  $C_M^k$ . 它还可以有不同的沟通模式,最简单的是单侧弱连通形式(如单条路径),需要  $k-1$  个连接,这些连接最少需要 1 个沟通模式实现(如果  $k-1$  等于  $n$ ),最多需要  $k-1$  个沟通模式实现;最复杂的是强连通形式(受树突与轴突数量的限制,当  $k > n$  时就不可能形成全连接形式),最少需要  $k-1$  个沟通模式(环),最多需要  $k \cdot C_n^1$  个沟通模式. 现在考虑,从确保能够建立连通结构的沟通模式数  $k-1$  开始,直到  $k \cdot C_n^1$ ,每多使用一个沟通模式就可视为一种新的连通结构,那么在数量级上连通结构形式的变化有  $k \cdot C_n^1 - (k-1)$  种. 因此,对于选择  $k$  个神经元的情形,有  $C_M^k \cdot [k \cdot C_n^1 - (k-1)]$  种不同的神经元连接模式. 这样,可能的组合数就是可能的最大容量: 
$$\sum_{k=2}^M \{C_M^k * [k * C_n^1 - (k - 1)]\} .$$

### 3.2 纠错分析

纠错包含由不完整刺激样本回忆和剔除多余的干扰刺激两个方面. 对不完整样本能否可靠地实现联想回忆取决于所对应连通结构的形式,如果是强连通的,则无论是样本哪一局部的信息出现都能完成整个样本的完全回忆. 所以,可以把问题转化成:刺激样本出现后,在给定的学习步内形成强连通的概率是多少. 假设有  $k$  个特征神经元被激活,每个神经元的输入野集合  $N_{fan\_in}$ /输出野集合  $N_{fan\_out}$  的分布都是随机形成的(因此,任意其他神经元被选作其输入或输出的概率是均等的),沿用以前的设定,这两个区域的尺寸为  $|N_{fan\_in}|=|N_{fan\_out}|=n$ ,它们在神经元总数为  $M$  的网络上扩散激励从而构成连通结构;同时,设定一个神经元活跃可能激活其输出野中  $1/L$  ( $0 < L < n$ ) 的其他神经元,时刻  $t$  每增加一次都会发生结构学习. 由于激励的传递像链式反应一样,所以,由一个活跃神经元激活的位于其输出野中的其他神经元可以激活更多的神经元,称为衍生输出野,与时刻  $t$  一起记为  $DeriveFanout^{(t)}(N_{fan\_out})$ ,所以在时刻区间  $[0, T]$  中活跃的衍生输出野是这样一个并集:  $DeriveFanout^{[0, T]} = \bigcup_{t=0}^T DeriveFanout^{(t)}(N_{fan\_out})$ ,此集合所表示的区域尺寸(在激励传导的动力学发生的过程中会有一些神经元失去活性,但只要刺激的不间断呈现,它们又会活跃起来. 活跃神经元的数量变化可以保持在一个可预测的范围内),  $|DeriveFanout^{[0, T]}| \leq \sum_{t=0}^T \left(\frac{n}{L}\right)^t \leq M$ , 增加的趋势呈上凸曲线型,趋近于  $M$ . 而且在有多个特征神经元被激活的情况下,由于存在输出野的相交,所以实际每个输出野中活跃神经元的数目要大于  $n/L$ . 求和时的偏大与加数本身的偏小可以平衡掉一些误差,因此近似地,在时刻区间  $[0, T]$  后,独立事件“任意一个其他神经元正好位于此衍生

输出野”的概率是  $p = \frac{\sum_{t=0}^T \left(\frac{n}{L}\right)^t}{M}$ . 特别强调,当输出野相交程度越来越大且学习步数越来越多的时候,

$\left| \bigcup_{t=0}^T DeriveFanout^{(t)}(N_{fan\_out}) \right|$  有趋于  $M$  的趋势,但非常难以定量估计,所以规定  $\sum_{t=0}^T \left(\frac{n}{L}\right)^t$  的值用  $M * e^{-\frac{1}{\sum_{t=0}^T \left(\frac{n}{L}\right)^t}}$  来代替.

当  $k$  个特征神经元被激活时,事件“形成单侧弱连通”等价于事件“ $k$  中任意神经元或是位于  $k$  中其他神经元的衍生输出野中或是其自身的衍生输出野包含  $k$  中其他神经元”,进一步等价于事件“ $k$  中不存在一个神经元同时不位于  $k$  中任何其他神经元的衍生输出野中且其自身的衍生输出野不包含  $k$  中任何其他神经元”,这个条件概率为  $1 - (1-p)^{k-1} (1-p)^{k-1}$ ; 当在实验中设定参数  $M=1000, L=5, n=20, T=4, k=10$  时,上述概率等于 99.945%. 更高的要求是,当  $k$  个特征神经元被激活后“形成强连通”,这个事件等价于事件“ $k$  中任意一个神经元均位于  $k$  中其他神经元的衍生输出野中且其自身的衍生输出野包含  $k$  中其他任意神经元”,这也是条件概率,为  $(p^{k-1} p^{k-1})^k$ ; 当我们设定参数  $M=1000, L=5, n=20, T=5, k=30$  时,此概率等于 27.95%, 当  $T=7$  时,此概率为 72.71%. 对比这两个概率可见,要得到一个一般的连通结构(仅“留下个印象”的记忆)是很容易实现的;而要得到一个强连通结构(要“刻骨铭心”的记忆)是不容易的,学习次数越多、参与的神经元越多就越容易实现,这在直观上很好解释.

两个有相同局部分支的连通结构可能不能被正常分开,当所呈现的刺激正好是以前学习的若干样本的共

同局部时,可能激活若干个有相似之处的知觉的神经表达模式,它们会产生干扰.出现这种情况是正常的,它是由外部世界在具有多样性的同时又具有相似性决定的.人的记忆经常会出现这种现象,就像看见轮子会使人联想到各种不同的车一样.

#### 4 进一步的讨论

从实验中可见,基于结构的表达决定了为了建立一个实现对知觉进行记忆的连通结构表达,其所需学习的次数稳定,不随样本尺度规模的变化而变化,而且快.这种快速性和稳定性是和网络能以很高的概率保证在一定的学习步数内实现单侧弱连通相一致的,前者是后者的体现;而且在增加学习步数和连通结构规模的情况下也能以稳定的概率得到强连通结构,收敛性好.

本文虽然讨论的起点是一种自联想记忆模型,但其实它还仅是直接表达方式的一个侧面,将其看作知觉信息的直接表达方式就能与许多问题发生密切的联系,因为知觉问题是其他认知问题的基础,是一个核心问题.对这个问题的研究使得可以在一个一致的认知结构模型的框架内讨论知识表示与概念化、认知过程是否可分开来研究、学习是否能和知识获取过程分离、思维过程是否为可语言化的等人工智能的最基本问题.将其应用到语义的直接表达上进行基于语义相关而不是基于符号匹配的推理方式的研究,这种推理方式所表现出来的是语义模式变化的演进.作为一种直接的表示方式对知觉单元进行表达,在神经连接形态上是统一的,从而超越了不同感觉通道的模态限制,可以使得图像、声音等信息在一个表达方式下表示.此联想记忆模型通过设定动力学方式、矩阵特征向量这两种方法来保持连通结构,是为了保持神经元的高度可重用性、非专一性.这同理于任何单个神经元所承担的语义都非常微小,沟通神经元甚至根本不承担语义,语义的呈现是大量神经元整合的结果,整合的方式设想为建立在有向连通结构上的动力学过程.

文献[26]对于概念语义的表达也有类似于本文直接表达的观点.它认为“语言的笔触就是思维的笔触”,高层次的意义组合和迁移同构于低层次的固定的神经过程,只不过同构已经变得极其复杂、微妙、精巧和多面化.如果按他的说法把神经表达看作“活跃的符号”,那么所有基于传统符号逻辑的形式化人工智能方法都能映射到神经表达上,这是很好的还原和回归,因此这些模拟智能的方法在哲学上就不再是用来解决问题的方法,而是用来描述问题的抽象,是对本质上由神经活动支配(neuron-bounded)的思维的宏观抽象描述.

本文所述的模型更多的是基于生物神经网络模型(BNN),而不是传统的人工神经网络模型(ANN).为了使对神经网络模型的研究更多地由“把 ANN 当作一般的并行分布式处理手段”转向“将其作为显现智能的物质基础”方面,我们应当更深入地了解智能的本质.

**致谢** 复旦大学脑科学研究中心的寿天德教授、顾凡及教授,复旦大学神经生物学研究所的长江学者、国家杰出青年基金获得者李葆明教授在神经信息处理与记忆研究方面给予本文很多指导,特此致谢.

#### References:

- [1] Thompson RF. Neurobiology of learning and memory. *Science*, 1986,233(4767):941~947.
- [2] Carew TJ. Molecular enhancement of memory formation. *Neuron*, 1996,16(1):5~8.
- [3] Kandel ER, Schwartz JH, Jessell TM. *Principles of Neural Science*. 4th ed., New York: McGraw-Hill Companies, Inc., 2000. 175~316.
- [4] Wilson RA, Keil FC. *The MIT Encyclopedia of the Cognitive Sciences*. Cambridge: MIT Press, 1999. 514~523.
- [5] Lepage M, Habib R, Cormier H, Houle S, McIntosh AR. Neural correlates of semantic associative encoding in episodic memory. *Cognitive Brain Research*, 2000,9(3):271~280.
- [6] Kandel ER. The molecular biology of memory storage: A dialogue between genes and synapses. *Science*, 2001,294(5544): 1030~1038.
- [7] Goldman-Rakic PS. Neurobiology of mental representation, In: Morowitz H, Singer J, eds. *The Mind, the Brain, and CAS, SFI Studies in the Sciences of Complexity, Vol XXII*. Reading: Addison-Wesley, 1995. 51~61.

- [8] Singer JL. Mental process and brain architecture: Confronting the complex adaptive systems of human thought (an overview). In: Morowitz H, Singer J, eds. *The Mind, the Brain, and CAS, SFI Studies in the Sciences of Complexity, Vol XXII*. Reading: Addison-Wesley, 1995. 1~29.
- [9] Sharkey NE. An oral history of neural networks. *Artificial Intelligence*, 2000,119(1-2):287~293.
- [10] Hattori M, Hagiwara M. Multimodule associative memory for many-to-many associations. *Neurocomputing*, 1998,19(1-3):99~119.
- [11] Sussner P. Observations on morphological associative memories and the kernel method. *Neurocomputing*, 2000,31(1-4):167~183.
- [12] Cheng ACC, Guan L. A combined evolution method for associative memory networks. *Neural Networks*, 1998,11(5):785~792.
- [13] Ritter GX, Diaz-de-Leon JL, Sussner P. Morphological bidirectional associative memories. *Neural Networks*, 1999,12(6):851~867.
- [14] Bosch H, Kurfess FJ. Information storage capacity of incompletely connected associative memories. *Neural Networks*, 1998,11(5):869~876.
- [15] Ferster D, Spruston N. Cracking the neuronal code. *Science*, 1995,270(5237):756~757.
- [16] Sakurai Y. How do cell assemblies encode information in the brain? *Neuroscience and Biobehavioral Reviews*, 1999,23(6):785~796.
- [17] Quinlan PT. Structural change and development in real and artificial neural networks. *Neural Networks*, 1998,11(4):577~599.
- [18] Bickhard MH, Terveen L. *Foundational Issues in Artificial Intelligence and Cognitive Science*. Amsterdam: Elsevier Publishing Company, 1995. 11~18.
- [19] Wermter S, Austin J, Willshaw D, Elshaw M. Towards novel neuroscience-inspired computing. In: Wermter S, Austin J, Willshaw D, eds. *Emergent Neural Computational Architectures Based on Neuroscience*. Berlin: Springer-Verlag, 2001. 1~19.
- [20] Tanaka K. Representation of visual features of objects in the intertemporal cortex. *Neural Networks*, 1996,9(8):1459~1475.
- [21] Fuster JM. Cortical dynamics of memory. *Int'l Journal of Psychophysiology*, 2000,35(2-3):155~164.
- [22] Sandler U, Tsitolovsky L. Fuzzy dynamics of brain activity. *Fuzzy Sets and Systems*, 2001,121(2):237~245.
- [23] Glassman RB. Hypothesized neural dynamics of working memory: Several chunks might be marked simultaneously by harmonic frequencies within an octave band of brain waves. *Brain Research Bulletin*, 1999,50(2):77~93.
- [24] Cariani P. Symbols and dynamics in the brain. *BioSystems*, 2001,60(1-3):59~83.
- [25] Rosser RA. *Cognitive Development: Psychological and Biological Perspectives*. Needham Heights: Simon & Schuster, Inc., 1994. 285~289.
- [26] Hofstadter DR. *Gödel, Escher, Bach—An Eternal Braid*. Beijing: Commercial Press, 1997. 456~459 (in Chinese).

#### 附中文参考文献:

- [26] Hofstadter DR. 哥德尔、艾舍尔、巴赫——集异璧之大成.北京:商务印书馆,1997.456~459.