



Thesis Submitted to Tianjin University of Technology for
the Master's Degree

**The research of pattern recognition
based on Bayesian network**

By
Lei Wang

Supervisor
Xian-bin Wen

January 2010

独创性声明

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作和取得的研究成果，除了文中特别加以标注和致谢之处外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得 天津理工大学 或其他教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文作者签名： 王磊 签字日期：2010 年 1 月 25 日

学位论文授权使用授权书

本学位论文作者完全了解 天津理工大学 有关保留、使用学位论文的规定。特授权 天津理工大学 可以将学位论文的全部或部分内容编入有关数据库进行检索，并采用影印、缩印或扫描等复制手段保存、汇编，以供查阅和借阅。同意学校向国家有关部门或机构送交论文的复本和电子文件。

(保密的学位论文在解密后适用本授权说明)

学位论文作者签名： 王磊

导师签名： 王磊

签字日期：2010 年 1 月 25 日

签字日期：2010 年 1 月 25 日

摘要

根据对哺乳动物大脑皮层的解剖学和生理学的研究，越来越多的证据表明虽然大脑皮层分为不同的功能区，但大脑皮层神经细胞和神经细胞的组织结构是相同的。大脑皮层在解剖学上分为 6 层，每个信息处理单元是一小片贯穿 6 层的细胞，称为神经柱（cortical columns）。处于不同功能区的神经柱可能使用同样算法处理信息；同时大脑皮层在逻辑上具有层级结构，信息在沿层级结构向上传递的过程中被归纳总结，不同层级的信息处理单元也使用同样的算法；而且研究表明连续时间信息在大脑皮层的学习和识别功能中起着重要作用。

基于对大脑皮层的研究 Jeff Hawkins 提出了记忆——预测模型，认为大脑皮层层级结构向上传递证据与存在于记忆中的证据相结合进行识别与推理，同时推理的结果作为证据向下传递修正记忆以实现预测。George Dileep 在记忆——预测模型的基础上提出了层级时间记忆（Hierarchical Temporal Memory）的数学模型，使用贝叶斯网络和信念传递实现证据的向上/向下传播。网络中的每个节点使用相同的算法，使用时间信息对证据进行分类。本文在上述工作的基础上，首先将层级时间记忆模型应用于图像识别，拓宽该模型的应用领域；然后使用几种聚类方法改进该模型的推理算法和学习算法，提出贝叶斯网络节点中的聚类算法。

关键词： 大脑皮层 贝叶斯网络 聚类 图像识别

Abstract

After many anatomical and physiological researches of mammalian neocortex, more and more evidences were found to prove that no matter what function the field of neocortex provides, the neurons and organizing of cortex cells are uniform. The neocortex can be divided into 6 layers anatomically. There're slices, each slice consist of several neurons through 6 layers is the unit of information processing, called "cortical columns". Cortical columns in different fields process information in same algorithm. Moreover, there's logical hierarchy in neocortex. Evidences are summary when passing up through the hierarchy. Different layers in hierarchy also use same algorithm. Besides, there're evidences proved that temporal information plays an important role in learning and inference of neocortex.

Base on these, Jeff Hawkins proposed the theory of memory-predice framework, considered that evidences are being passed up in hierarchy, combined with evidences in memory to achieve inference, meanwhile, evidences are being passed down to revise the evidences in memory to provide predice. George Dileep proposed mathematical model called Hierarchical Temporal Memory framework base on memory-predict framework. It has the topology of bayesian network and using belief propagation to pass evidences in hierarchy. Every node in network has same algorithm, and using temporal information to categorize the evidences. In this thesis, base on the work mentioned above, we first implemented this framework in visual category recognition field, broaden the application area of this framework, and then we replaced the cluster algorithm in nodes with several other cluster algorithms to improve to framework, propose the appropriate clustering algorithm in the nodes of Bayesian network.

Key words: Mammalian neocortex, Bayesian network, clustering, recognition,

目 录

第一章 绪论	1
1.1 研究目的和意义	1
1.2 国内外研究现状分析	1
1.3 目前国内外研究所存在的问题	3
1.4 本文的研究内容和组织	4
第二章 大脑皮层的生理学基础与贝叶斯网络	5
2.1 研究大脑皮层的意义	5
2.2 大脑皮层和神经细胞	7
2.2.1 大脑皮层和神经细胞结构	7
2.2.2 神经系统的调节	9
2.2.3 神经柱	9
2.2.4 大脑皮层中的时间与空间层级结构	10
2.2.5 皮质微通路	11
2.3 大脑皮层的运作理论	12
2.3.1 记忆—预测模型	12
2.3.2 时间信息在大脑学习和识别中的作用	12
2.4 贝叶斯网络	16
2.4.1 贝叶斯网络定义	16
2.4.2 贝叶斯网络图形表示	17
2.4.3 贝叶斯公式和规则	17
第三章 层级时间记忆模型	19
3.1 HTM 网络的拓扑结构	19
3.2 新大脑皮层层级结构与 HTM 网络层级结构的映射	20
3.3 节点的运作方式	21
3.3.1 节点的学习	22
3.3.2 节点的感知和推理	26
3.3.3 层次结构中节点的操作	26

3.4 噪声处理	28
3.4.1 有噪声输入的学习	28
3.4.2 有噪声输入的感知/推理	29
3.5 HTM 网络中的信念传播	30
3.5.1 动态情况下的信念传播公式	35
3.5.2 HTM 网络中的反馈传播	37
3.5.3 HTM 节点中的信念传播公式	37
第四章 基于时间层级记忆模型的图像识别	40
4.1 对 USPS 数据集的识别	40
4.2 对 ALOI 数据集的识别	42
4.3 本章小结	43
第五章 时间层级记忆模型中的改进聚类方法	44
5.1 时间层级记忆模型中使用的聚类方法	44
5.1.1 κ 均值方法	44
5.1.2 分层聚类算法	45
5.2 对输入模式的聚类算法进行改进	45
5.2.1 最简单聚类方法	45
5.2.2 最大距离样本法	46
5.3 对于按时间分组过程中聚类算法的改进	47
5.3.1 K 均值方法	47
5.3.2 叠代自组织 (ISODATA) 聚类法	48
5.4 本章小结	50
第六章 结束语	51
6.1 本文主要工作	51
6.2 有待进一步研究的内容	51
参考文献	52
发表论文和科研情况说明	55
致 谢	56

第一章 绪论

1.1 研究目的和意义

随着计算机技术与信息技术的发展,图像识别技术获得了越来越广泛的应用。图像识别技术的涵义很广,主要指通过计算机,采用数学技术方法,对一个系统前端获取的图像按照特定目的进行相应的处理。图像识别包括诸如生物特征识别(人脸识别、指纹识别等)技术、智能交通中的动态对象识别、手写识别等。可以说,图像识别技术就是人类视觉认知的延伸,是人工智能的一个重要领域,随着计算机技术及人工智能技术的发展,图像识别技术越来越成为人工智能的基础技术。它涉及的技术领域也越来越广泛,应用越来越深入。其基本分析方法也随着数学工具的不断进步而不断发展^[1]。

在图像识别技术中,从识别逻辑的观点来看可分为两个类型:组合式的和顺序式的。前者是把图像的特征全部抽出(或足以判别一个图像的很大一部分特征)之后再行判断,给出结果。后者则按所抽出特征的次序,每抽出一特征,都要进行一次判断(不是对整个图像),直到最终给出结果。除了这两类识别外,近年来模糊数学和神经网络在图像识别领域中也得到了广泛应用^[2]。

贝叶斯网络是R. Howard和J. Matheson于1981年提出,80年代早期,成功地应用在专家系统中对不确定性知识的表达,80年代后期,贝叶斯网络得到迅速发展,90年代后,研究人员开始尝试直接从数据中学习并生成贝叶斯网络的方法,取得了可观的成果^[3]。本课题的研究目标是通过设计基于贝叶斯信度传播网络的图像识别框架来实现图像的分类与识别。

1.2 国内外研究现状分析

贝叶斯网络(也称为信度网络、因果网络或者推理网络)是一组概率变量、节点和其间的有向连接。每一节点有一组有限的排他状态,其似然分配表示为信度值。有向连接表示它们之间的相关或者推论关系。事件本身的不确定性以节点的信度值表示,专家知识的确定性以条件概率表示。通过在贝叶斯网络中输入证据,更新节点的后验信度值,则可以对节点事件进行推理^[4]。贝叶斯网络目前广泛应用于辅助智能决策、数据融合、医疗诊断、文本理解以及数据挖掘等领域。在计算机视觉领域,贝叶斯网络已经有了广泛的应用,如目标检测、目标追踪、视频监控、目标识别等^[5]。

贝叶斯学习也是一种重要的机器学习算法,不确定性知识的表达和推理长期以来一直被认为是人工智能领域中的一个重要研究问题,在现有的方法中,概率方法是最为流行的方法,贝叶斯推理提供了推理的一种概率手段,贝叶斯网络为具有不确定性领域的

建模提供了一种非常有效的工具，它是结合概率理论和图论的一种不确定性知识的表达和推理的模型，表现为一个赋值的复杂因果关系网络图，它主要由两部分组成：

(1) 有向无环图(directed acyclic graph, DAG)，即网络结构，包括节点集和它们之间的连接。图中每个节点表示一个有一组状态的随机变量，各随机变量之间的弧表示事件发生的直接因果关系；

(2) 反映节点之间关联程度的概率参数，即条件概率表(conditional probability table, CPT)，节点通过条件概率表与同它相连的节点关联，并定量地表示关联的程度。它利用图形模式描述变量集合间的条件独立性，而且允许将变量间依赖关系的先验知识和观察数据相结合。

混合贝叶斯网络(hybrid Bayesian networks, HBN)就是既含离散节点也含连续节点的贝叶斯网络。

模式识别是指运用分析算法进行分类和识别^[6]。它诞生于20世纪20年代，随着40年代计算机出现和50年代人工智能技术的兴起，在60年代，模式识别迅速发展成为一门新的学科。模式识别所研究的理论和方法在很多科学和技术领域中得到了广泛的重视，拓宽了计算机的应用领域。近几年来，模式识别研究取得了重要成果，在工业、农业、国防、科研、医疗卫生、气象、天文等许多领域得到了成功应用。

图像识别作为模式识别的一个分支在研究过程中沿用了模式识别的关键环节。目前国内外的研究中，一般根据预定义的数据对图像提取关键特征，再使用不同的分类器对特征的属性数据进行分类，从而实现分类或识别目标。使用贝叶斯网络方法实现的图像识别即是在分类过程中使用贝叶斯网络作为分类器，使用训练数据训练得到先验概率表。在推理过程中再使用贝叶斯公式得到后验概率，选择概率最大的作为目标的类别^{[7]-[12]}。

事实证明这种方法在一定程度上是可行的，针对特定情境下的特定目标能够提供较高的识别率，并且已经广泛应用在生物特征识别(人脸识别、指纹识别等)技术、医疗图像识别、手写识别等领域。但是这一类方法存在的关键问题是特征属性的提取是针对特定情境下的特定目标的预先定义在系统中的局部特征的提取(如眼睛^[12]，玉米叶病害^[8]，电厂^[10]等)，这样一个系统建立后只能识别一类图像而无法通过不同的训练数据识别不同类别的图像。所以这一类方法缺乏普遍性，同时也不是人脑进行视觉识别的方法。而且由于我们无法为现实世界中的每一类物体都建立一个这样的识别系统，而人却可以通过学习识别每一种学习过的物体(即使经过了严重的形变)，所以可预计未来的图像识别以及智能计算技术将朝着仿照人脑的运行方法的道路发展。

近几年，针对人脑运行机制的研究在神经医学领域和智能计算领域都有了很大的发展。其中，J. Hawkins在总结前人对于人脑的生理学研究的基础上，提出了一种记忆——预测模型，认为人是通过记忆在脑中的信息进行预测，再对比感受器官所感受到的信息来进行判断和控制自身行为的^[13]。这种理论的数学模型接近于贝叶斯信度传播网络^{[14][15]}，记忆的信息为不同类别的模式及对应的一系列先验概率。通过计算得到观测值的

概率,即为预测。再用实际观测到的后验概率计算信度来更新网络中的先验概率,从而使记忆更接近于现实。现代医学的研究成果证实大脑皮层的神经细胞是成层级式连接的(其中视觉区域的脑皮层细胞有4个层次,分别是V1, V2, V4, IT层,细胞数量依次由多到少,单位识别区域由小到大)^{[13][16][17]},所以有学者提出了层级贝叶斯网络模型来模拟人脑皮层中视觉区域神经细胞的行为进行图像识别^{[18][19]}。

存储在脑中的事物是通过恒定表征实现的,这样才能实现在不同光照条件、不同观察角度、物体变形、噪声干扰等条件下仍能够保证对物体的识别,这就需要提取物体的特征。但有时同一事物在不同状况下的表征差别很大(例如一只狗的正面和侧面,或是同一个人的脸在正面光照和被阴影遮住一半的情况下)。要将这些识别出恒定表征难度较大,一般方法是寻找在变化中不变的局部特征进行提取,这就导致了要想实现类似于人脑的全面的人工智能,就必须对每种事物的每一类变化在系统中定义要提取的特征,这在现实中是不可行的。并且医学研究发现虽然大脑皮层分为运动区、视觉区、听觉区等不同的功能区域,但是不同区域的脑皮层细胞内部结构和细胞的组织方式都是相同的,这说明人脑是通过一种统一的算法来实现恒定表征的识别的(甚至是通过一种一般化的算法来处理各种感官信息和控制行为的)^[13]。

对此, J. Hawkins认为时间在其中扮演着重要的角色,在参考前人对使用时间作为参量进行识别的研究的基础上^{[20][21][22]},提出在临近时间上观察到的模式因为很可能是由于现实世界中的同一事物所致所以应当依概率被归为同一组^{[13][23]}。

据此理论框架 Dileep George 与 Jeff Hawkins 提出一种基于贝叶斯信度传播网络的模型,称为层级时间记忆模型(HTM, Hierarchical Temporal Memory)^{[15][19][24][25]}。使用 HTM 模型进行推理时,由于被识别图像具有噪声等原因,输入信息被转化为空间模式的概率分布,与先验概率结合作为信度随层级结构向上传播,最终得到输入图像属于某一类别的概率信息。在向上传播的过程中算法为简化将同一层级的不同节点认为是独立的,采用朴素贝叶斯的方式,在上层节点简单计算其子节点的概率之积^[26]。

1.3 目前国内外研究所存在的问题

如前所述,普通贝叶斯图像识别方法存在特征属性提取是针对特定情境下的特定目标的预先定义在系统中的局部特征的提取,系统建立后只能识别一类图像而无法通过不同的训练数据识别不同类别的图像,缺乏普遍性和扩展性,同时对于形变较严重的目标识别率低,鲁棒性较差。

而记忆一预测模型模仿大脑皮层的运行机制,具有很强的神经生理学研究成果基础,又具有贝叶斯网络作为其数学基础,使其理论具有很强的健壮性。同时,从 HTM 模型的原型程序可以看到其已经初步实现了许多其他方法没有实现的功能,但是几处关键

算法采用了简化的方法，例如假设同一层节点间相互独立（但实际上脑皮层细胞同一层级之间有非常多的连接，证明其互相有非常紧密的联系^[13]）；使用时间邻接矩阵将空间模式分组时使用贪婪算法，其为近似算法并不一定能够达到最优；此外，HTM 模型包含了上行和下行两条信度传播途径，但原型程序为了简化只实现了上行^[25]。其中某些原因导致了系统的识别率较低，某些人能够明显识别的图像系统识别出错。

另一方面，由于会在较低层保存大量的模式，并存在着大量的矩阵相乘、矩阵与向量相乘的运算，使得系统的性能较低，训练和识别的时间延长^{[28][29][30]}。

1.4 本文的研究内容和组织

记忆—预测模型模仿大脑皮层的运行机制，具有很强的神经生理学研究成果基础，又具有贝叶斯网络作为其数学基础，使其理论具有很强的健壮性。因此本论文将沿用这一理论基础，根据记忆——预测模型理论，设计一种基于贝叶斯网络的图像识别框架，对不同种类物体采用统一的训练识别方法，在具备识别严重形变物体类别，及识别同类的不同物体的基础上，提供较高的识别率和较好的计算性能。

本课题的研究目标就是把现在的算法进行改进，简化不必要的或低效的计算，加强关键部分算法的精度，以期能够在提高系统精度的同时提高系统效率或者使二者更好的平衡以达到实际应用的目的。

因此这里要解决的关键问题有：

- (1) 通过寻找适合的阈值参数降低系统计算量；
- (2) 寻找全局最优的图分割算法处理空间模式分组的问题；

第二章 大脑皮层的生理学基础与贝叶斯网络

2.1 研究大脑皮层的意义

了解新大脑皮层的运行方式是制造智能机器的必不可少的一步。可以从机器学习的方式来讨论研究大脑的必要性。在机器学习中，研究者首先针对需要建模特定数据构建一个参数化适应模型。初始模型基于领域知识构建，包含有能够根据训练数据调节的参数。而后研究者设计出能够调整这些参数的算法。通常参数调整的目标是使错误最小化^[31, 32]。参数调整后的模型被用来为新的模式分类，预测未来的模式，压缩，重建等。但是这种方法遮掩了一些与学习相关的基本问题。

一个方面是学习的样本的复杂性，即为模型的性能能够达到一定水平所需的训练样本。如果一个学习问题的假设空间是非常大的，那么构建一个完成学习的模型的需要大量的训练数据和较长的训练时间。我们可以看到，面对不同的学习问题时，没有一种学习算法的优越性全面超越其他所有的算法。如果一个算法对某一特定问题优越，只是因为利用该算法假设，这一问题合适的。同样的算法将不好的一个问题，已经假定从原来的不同。这意味着，使有效的算法，机器学习研究人员对编码到最初的模型结构的问题领域知识。先验知识越多，就是把，越容易学习。某一算法在解决特定问题时的优越性是因为设计这种算法所基于的假设原理更加适合于这一类问题。而在面对不适合的问题时，这一算法就可能无法表现出优越性。这意味着为了设计出有效的算法，机器学习研究人员需要将一部分涉及到领域问题的领域知识设计到初始状态模型的结构或算法当中去。而在初始状态模型中置入的先验的领域知识越丰富，模型就会越容易训练并且更通常能达到更好的性能。但是我们却无法做到为每一个需要解决的问题设计一种单独的机器学习模型。

另一方面，人类和其他哺乳动物似乎在用一种完全不同的方式来解决这些问题。人类可以学习和适应初始状态模型（大脑皮层）成立时并不存在的问题这一事实，是由人类大脑采用这种通用机制的证明。此外，大量的现在研究使研究人员推测新大脑皮层在学习不同问题时使用一种基本相同的算法。也就是说在面对不同的学习模式的问题，如听觉、视觉、体感知觉和语言时使用了同一种算法。

虽然没有被最后证明，但是越来越多的证据证明了一种通用大脑皮层算法的存在。存在一个通用的皮层算法的推测首先由一位研究了不同哺乳动物皮层组织的神经科学家提出^[33]。不同功能区之间，甚至在不同的物种之间皮层连接方式的显著的相似性构成了这一猜想的基础。对雪貂的实验结果为这一理论提供了进一步的支持^[34]。在实验中，

雪貂的视觉神经纤维被连接到了大脑皮层的听觉区域。经过一段时间视觉信号的刺激，雪貂的听觉区发展出了视觉感受区，并能够借此进行日常行为。

从这两方面分析可以看出如果我们设计出一种通用的机器学习的模型和算法来解决各类问题，如视觉、听觉和语言问题，而大脑皮层可以使用一种通用算法来解决各类问题，所以找到大脑皮层在处理问题时采用的通用方法是非常关键的。而从另一方面看，大脑皮层能够以用一种通用机制来解决类型广泛的问题，那么我们也可以肯定在看似不同的各种问题中应该存在一种共性。而这一共性是与大脑皮层解决问题的机制相契合的。

建立一个通用学习理论的关键是一个通用的建设理论。学习算法的归纳偏置是指模型在遇到一个未曾见过的输入后使用假设集来预测输出^[35]。我们预先设定的假设越多，模型学习起来越容易，但是同时模型能够解决的问题越少。所以如果要设计一种能够解决各种问题的算法，那么问题的关键就在于如何找到最小的假设集来保证模型学习的能力和解决问题的广泛性之间的平衡。所以我们需要研究大脑皮层来找到存在于我们大脑中的关于周围世界的假设集。

仿制大脑皮层的逆向工程是非常困难的，我们无法确定这种通用假设集。此外我们也无法确定关于大脑的众多解剖学和生理学的研究中的哪些是至关重要的，哪些我们可以从中发掘出这种通用算法。

一种较好的策略是同时研究大脑皮层与周围世界，对于大脑皮层的解剖学和生理学的研究可能会为我们提供大脑自然产生假设的重要线索，通过研究脑皮层的组织结构，我们可以以学习的视角找到与机器学习相关的一般规则，再从中选取与周围世界相关的部分。如果内部脑皮层的组织形式与外部世界的组织形式相吻合，那么就可以认定我们找到了需要的假设规则。

2.2 大脑皮层和神经细胞

2.2.1 大脑皮层和神经细胞结构

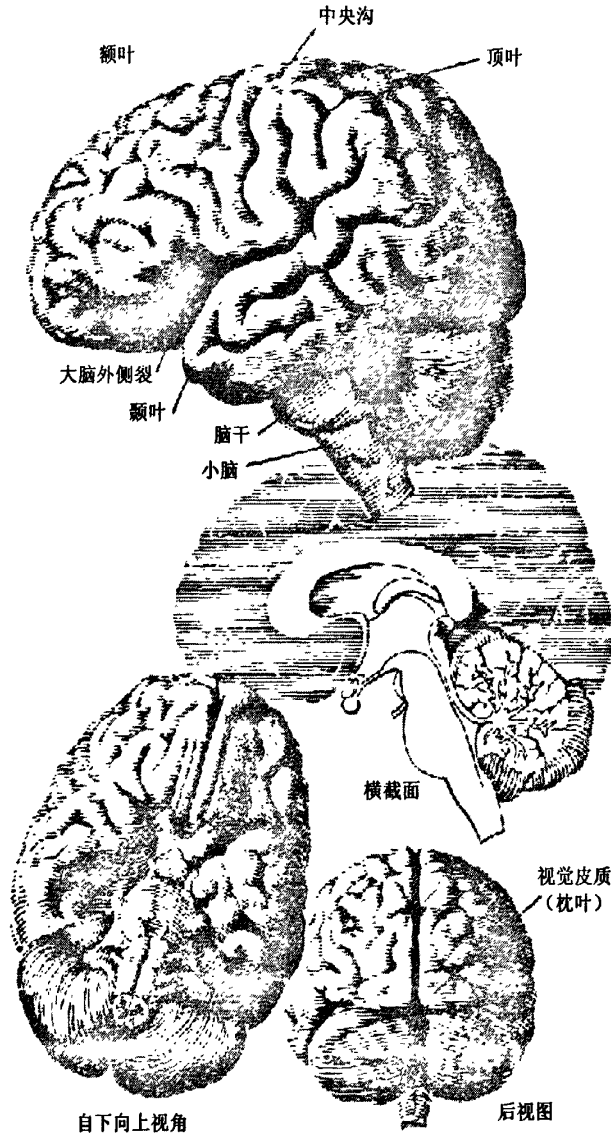


图 2-1 大脑皮层解剖图^[36]

单个的神经细胞呈现一种树形结构，其形似一条主干并在其上分离出很多树枝。神经细胞中的主干部分称作轴突。神经冲动以一种电信号的方式在神经细胞内传播，神经冲动的电脉冲只有十分之一伏，并且只能持续约千分之一秒，电脉冲在轴突内的传播速度能达到 500 公里/小时。

单个的神经细胞呈现一种树形结构，其形似一条主干并在其上分离出很多树枝。神经细胞中的主干部分称作轴突。神经冲动以一种电信号的方式在神经细胞内传播，神经

冲动的电脉冲只有十分之一伏，并且只能持续约千分之一秒，电脉冲在轴突内的传播速度能达到 500 公里/小时。

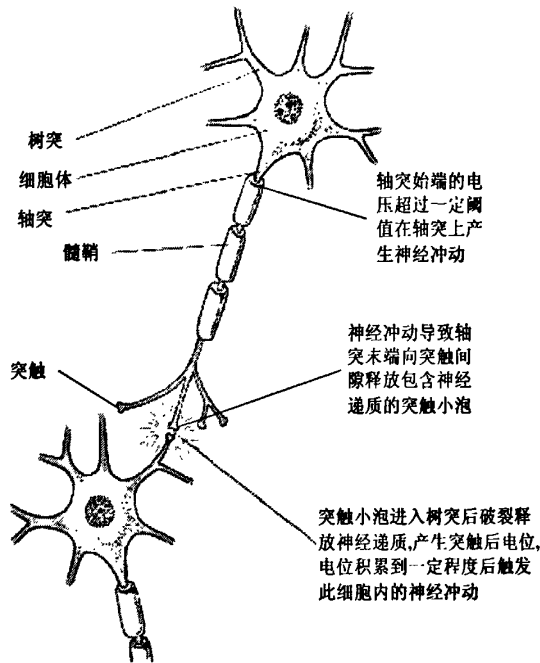


图 2-2 神经细胞结构

从整体上来说神经冲动是以一种接力赛的形式传播的。相邻的神经细胞的轴突与树突相连，组成一种细胞链形式，而大脑就是通过这种神经细胞链最终连接到各种感官的感受器上的。神经冲动传播时，轴突中的电脉冲信号发生瞬间电压变化触发其末端突触从而向与之连接的另一神经细胞的树突释放一种称为神经递质的化学物质（主要组成成分包含乙酰胆碱、 γ 氨基丁酸等）。

接收到神经递质的细胞会在其内部产生电压变化，但这一变化并不能简单造成神经冲动的继续传播。由上图可以看到每一神经细胞有一个轴突和许多个树突，这意味着单个神经细胞可能会同时接收到许多输入信号，并且输入信号还包括正电压（刺激信号）和负电压（抑制信号）两类。神经细胞如同计算器般累积的所有输入的正、负电压之后，所得的最终电压是正的并且达到一定阈值后才会触发神经冲动在细胞中的继续传播。此外输入信号的累积电压还会决定神经冲动的脉冲频率（脉冲频率在 0-1300 次/秒之间变化）。

大脑皮层厚度只有几毫米，一般认为可以分为 6 层，分别用 I-VI 表示，一些层也可以细分为子层，如 IVc 层。较大的神经细胞一般存在于更深的层，如层 V。从感觉习惯输入的信号一般被发送到固定的层次，例如由眼输入的信号被输送到视觉皮质的 IVc 层。而接近于大脑表面的 I, II, III 层目前还没有被充分的了解。

2.2.2 神经系统的调节

网状激动系统具有调节输入大脑的信息流的能力，这是通过一系列的下行连接实现的。这种下行连接能够通过对于感觉输入中的异常调节感觉系统的敏感性^[36]。这种机制的好处在于帮助大脑在连续不断的刺激中判断重要性事物，例如在大脑持续地接收皮肤同衣物接触的感觉信号的同时能够及时判断出蚊子落在身上产生的一种新的感觉。这种功能在生物的生存和进化中的功能是相当重要的。

网状激动系统也以同样的方式调节对于疼痛感觉的输入。所以，治疗剧痛的一种方式就是通过改变网状激动系统中对于疼痛的敏感性来达到减弱疼痛的目的。从重复感觉输入中辨别异常的能力是非常重要的，所以网状激动系统不是神经系统中唯一负责此项功能的机制。弱化重复感觉的机制存在于神经系统许多层次中，从最低的单个神经细胞的层次此机制就开始存在。当神经冲动在神经细胞间传播时，发送方释放一定量的神经递质到接收方，经过一段长时间的神经冲动刺激后，每次神经冲动脉冲所释放的神经递质的量会产生变化，在某些情况下升高，而在另外的情况下会降低。对于一种称为“雨虎”的海参的神经系统的研究为这种调节机制提供了很好的例子。它具有一种简单的回避反射，当它腮周围的皮肤被触摸时就会收回，但是当这种触摸以每12秒一次重复进行时，它停止继续大幅度地收回腮的动作。这种对于持续刺激的反应的逐渐减弱被称为“习惯化”。对于皮肤刺激的效应减弱的原因在于经过一定数量的刺激后，传感神经细胞的轴突末端每次脉冲所释放的神经递质越来越少。但同时当遇到异常刺激时会恢复原先状态，在本例中如果在对腮部的持续刺激导致“习惯化”之后触摸“雨虎”的头部，它会恢复强烈的反射状态。

2.2.3 神经柱

首先介绍新大脑皮层的解剖学特性。新大脑皮层是人类和其它哺乳动物智能的基础。它看起来像一张薄片，只有几毫米厚度，包含有多层神经细胞。感知和运动过程在新皮层中是按区域组织的，这意味着邻近的神经元通常负责相关的功能。例如，视觉区V1皮质区中的神经元按特定结构排列，相邻神经元对应视网膜上的相邻接受区。不同的区域对应于不同的功能形式或信息处理的不同阶段^[37]。

皮层区域具有薄片状的组织形式，也就是说新大脑皮层是由许多薄层组成的，特定薄层的神经元符合规定了其从何处接受连接及向何处连接的正规模式。一般情况下认为新大脑皮层是由6个不同的薄层组成的^[38]。这些薄层被赋予了1到6的编号。层1是最接近头骨最远离白质的一层，层6最接近白质。在水平方向上，层1位于最高层，层6位于最底层，层1-3经常被称为表面层，层5和6被称为深层。

除了水平方向，皮质区域还具有垂直方向的组织，称为神经柱^[39]。这种垂直组织的特征是同一垂直柱中不同薄层的神经元具有高度的共性。同一柱中的神经元具有高度的

解剖学上的连通性。同一柱中的神经元还反映了生理学的属性。基本的视皮层神经柱无法识别形状的边缘。在更高层级的皮层区域如IT，单个的神经柱能够识别整个物体的视觉特征。虽然柱状的组织形式被广泛认同，但是它的功能还有争论。

现在被广泛认可的是不同的视皮层区域是以层级的形式连接起来的^[40]。例如基本视皮层，称为V1，是第一步处理视觉信号的皮层区，V1区处理的结果被发送到V2区之后是皮层的V4区。不同皮层区域间的连接穿过白质或皮质本身。并且皮质不同区域间的连接是相互的。事实上，视皮层不同区域在层级中的位置是基于前馈和反馈传播映射的模式在解剖学上确立的。

2.2.4 大脑皮层中的时间与空间层级结构

自然和人造的动力系统常常具有嵌套的多尺度组织结构。在这样的系统中，同时具有大尺度系统级的变化和小尺度子系统级的变化。通常大尺度系统级的变化相较于小尺度子系统级变化得更慢。

空间和时间都是按层级组织的，空间上的层级可以被理解为由子系统构建复杂系统，子系统提供了稳定且可重用的部分并通过装配组成了大系统。Herbert Simon是首先指出这种组织形式的人之一^[41]。他指出时间上的层级可以认为是子系统以特征频率组织的。不同的子系统因动力学特征具有不同的特征频率。通常情况下空间组织和时间组织也有所对应，空间尺度较小的子系统通常具有较高的频率，而较大的子系统具有较低的频率。

层级系统的不同层级之间具有双向的互动，较高层级的状态变量不是简单的低层状态变量的平均。低层的变量决定了高层变量的变动的频率，而高层变量为低层变量提供了参考信息。

可以确定的是视觉系统的组织是具有解剖学上的层级结构的^[40]。视觉皮层的层级组织结构是经过充分研究的，视觉信息离开视网膜后通过外侧膝状体核首先进入V1皮层区^[42]。在腹侧通路信息通常被认为是对于物体识别具有意义的，随后被继续传送到V2, V4和IT皮层区。V1区的神经细胞只能识别视野内的一小部分，V2区的每个神经细胞从几个V1区的细胞中接收信息。所以一个V2区细胞的有效接受范围几倍于V1区的单个细胞。随着信息的沿层级向上传播，单个细胞的有效接受范围越来越大。一个IT区的神经细胞的接受范围几乎相当于整个视野范围。当我们观察一个物体时，一个V1细胞可能只能观察其中的一小部分，如一条边线，而一个IT则能观察到整个物体。

视觉皮层内的信息流不仅由下至上传播。在不同层级间前馈和反馈连接同时存在。不同层级间的互动以周期性的方式发生，较高层次为较低层次提供参考信息。

与视觉皮层的层级结构同时存在的是时间层级。当我们在观察的同时慢慢移动物体，V1区的神经细胞会发现局部特征出现在接受范围随后又离开，而IT细胞能够一直观察到整个物体，这说明在较低层次特征变化较快而在高层次保持稳定，也就是说随着层

级的上升时间特性减慢。

The No Free Lunch (NFL)理论^[43]提出对于所有的学习问题没有一种学习算法比其他学习算法具有内在优势。如果一种算法针对某一特定问题具有优势，那么只能是在算法中加入了适合这一问题的假设。另一方面，神经科学的研究成果表明人脑皮层很可能使用同一种算法来解决诸如视觉、听觉等各种不同的问题。同时人脑皮层也是具有层级结构的。对于这一现象的一种合理解释是来自不同领域和感觉样式的信息虽然表面上千差万别，内在却可能具有同一种统计学属性。这可能是由于外部世界的信息生成机制因物理定律和白组织规则而具有内在的层级造成的。人脑皮层可能通过进化发现了这一事实并利用它为层级结构的数据建立了有效的模型。

2.2.5 皮质微通路

皮质神经元按不同的组织方式形成了由垂直神经柱和水平薄层构成的微通路连接。这些微通路的许多属性是固定的并且不随年龄、皮质区域、物种而变化。这说明神经微通路是由一般微电路模板演变而来的。计算神经科学的一个重要难题就是理解固有皮质微通路中特定薄层突触连接的计算功能。

关于薄层内部和薄层之间神经元的连接有许多解剖学和生理学上的研究。许多研究者通过整合大量解剖学数据提出了关于这些连接的图表。图2-3展示了其中的一种^[44]。在这一通路中，兴奋的前馈路径是通过如图所示轴突的投射模式这一解剖学证据得到的。

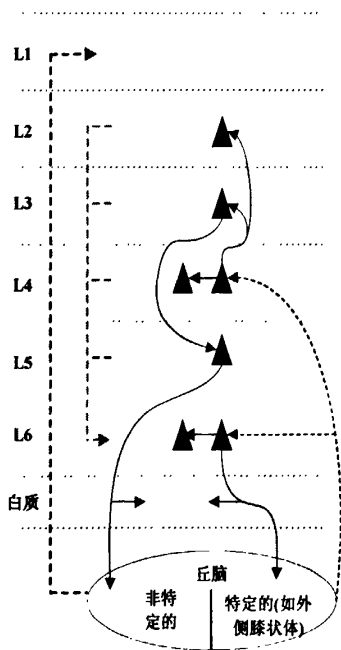


图2-3 皮质微通路一个示例图^[44]。

2.3 大脑皮层的运作理论

2.3.1 记忆—预测模型

Jeff. Hawkins在他的On Intelligence^[63]一书中依据生物学中的研究成果提出了一种解释大脑皮层运行方式的理论，称作记忆—预测结构。这一结构具有以下特征：

1. 大脑皮层为其所接触到的时空模式构建一个模型，目的在于通过这一模型为输入模式的下一步进行预测。

2. 大脑皮层的构建是通过复制一种被称为标准皮层回路的基础计算单元完成的。从计算机的角度，这种计算单元可以被看作重复出现的节点。

3. 脑皮层是按一种层级结构组织的，这意味着这些计算单元，即节点，是以一种树形结构连接在一起的。

4. 脑皮层通过在节点中记忆模式并序列化建立一种层级化的时空模式最终为皮层所接触到的外部世界建立一个模型，这一模型被用来对输入模式进行预测。

5. 大脑皮层建立外部世界模型是一种无监督学习。

6. 每个节点储存了大量的模式序列，大脑皮层就是使用这些模式序列进行模式识别的。

7. 节点所学习到的模式序列决定了该节点的输出。

8. 在层级结构中信息被向上或向下传递来实现对信息的识别和消除信息的模糊性。

2.3.2 时间信息在大脑学习和识别中的作用

视觉是人和哺乳动物感受世界的最基本感觉形式。与视觉处理相关的区域占了人类大脑皮层总数的30%。由物体反射的光线进入我们的眼睛在视网膜上形成一个像，我们的大脑接着通过解释这些图像来感受外部世界。虽然这看上去是自动切毫不费力的，但是在这一过程中我们的大脑解决了许多计算机至今未能解决的问题。

视觉的不变模式识别就是其中的一个问题。人类和哺乳动物在位置、大小、光线条件、形状变化的情况下以及存在大量噪声的情况下都可以做到识别。对于这些问题在计算机上进行过一些尝试，其中许多把不变模式识别问题归为分类问题。在大多数情况下，分类问题有固定种类数量的样本，赋予其中的每个种类一个标签，每个种类中包含了此类物体位移、变形、大小变化的情况。通过使用这些样本对分类器进行训练，使其能实现对于样本类别的物体能充分识别。

通过对于人类和其他哺乳动物处理视觉问题的方式进行分析可以发现，许多早期研究忽视了时间在视觉问题上的重要性。人类可以通过静止图片识别物体是在研究中忽视视觉的时间部分的主要原因。虽然人类能够识别静止图片，但是在我们的学习认知该物体时是通过持续变化的观察实现的并且通过时间信息我们可以学习认识到物体重要的共

有特征。哺乳动物学习的另一重要方面是无监督学习的性质。动物并不会为物体分类做标签。

为不同的类别作标签并加以训练的分类问题称为有监督学习。视觉模式识别问题一般按有监督学习问题处理，这可能是因为人类为世界上的每种物体都使用标签命名了的缘故。然而其它哺乳动物进行不变视觉模式识别的学习过程显示了视觉问题的学习是一种无监督学习。

例如一只哺乳动物朝着它食物走去，它每走一步，就会有一个食物的不同的像映到它视网膜上。两幅连续映在它的视网膜上的图像的欧几里得距离可能会很大，然而它仍然知道他在看着同一食物。动物无法为食物命名，也没有人教它食物叫什么，它对于事物的识别完全是一种无监督学习的过程。

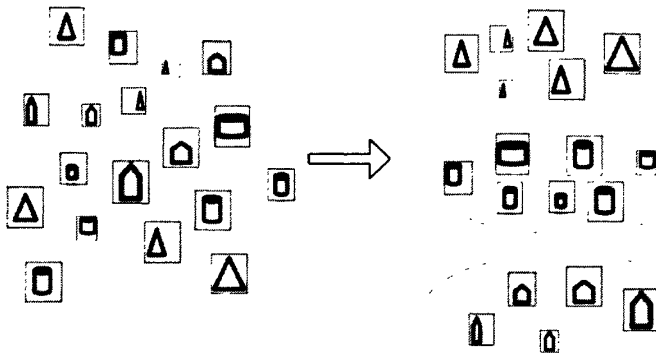


图2-4 不同物体的识别分类^[16]

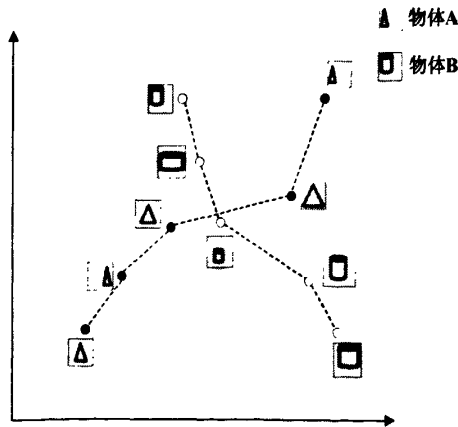


图2-5 分类后的两类物体^[16]

如图2-4所示，左侧是不同种类的物体，包括三角、圆柱形和类似房子的形状，那么如何使用无监督学习把这些物体像图2-4右侧部分那样分成三类：简单的像素尺度的度量无法实现这一目标，因为很多非同类物体的像素空间的重叠大于同类物体。属于同一类物体的不同图像只是形状和大小发生了变化。假设哺乳动物的视网膜上的像素数量为 N ，则输入的图像可以被认为是一个 N 维向量。由物体产生的不同图像对应于 N 维向量

空间的不同的点。属于同一物体的点集可以被看作是 N 为空间中的一个数集。这些数集可以包含无数的点，因为它们包括同一物体不同图像中的细微变化。不同的物体属于不同的数集。在动物的视网膜上形成的碗的图像可以被认为是“食物”这一数集中的样本。如图2-5所示

在图2-5中，有两个数集，对应于物体A和B。数集A中的不同点是物体A所产生的不同图像，数集B中的不同点是物体B所产生的不同图像。如果要正确地识别物体A和B所产生的图像，那么它必须能够使用无监督的方式分离这两个数集。

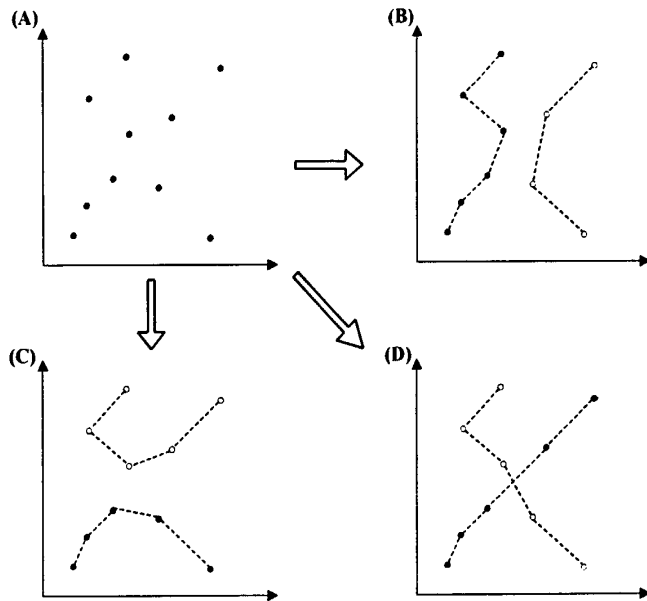


图2-6 三种不同的数集分裂方式^[46]

无监督学习要求在一个高维空间分离数集。给定数集上未标记的属于不同物体的点，有几种方式将它们分裂为不同的数集。图(A)所示一个高维空间中的未标记点。图(B)到(D)说明了几种将这些点分裂为两个数集的方式。

使用无监督方式分离数集的难点在于不同物体的数集相互交叉或过于接近。例如，物体A和B从某些角度看来是一样的。那么这些点就属于这两数集共有。如果两个数集的点有交叉并且点没有使用标签标注，那么在没有附加信息的情况下就无法分离这些数集。图2-6中显示了交叉数集的三种分割。所以如果我们将图像作为 N 为空间中的点来处理，那么我们就无法用无监督方式分离物体的数集。

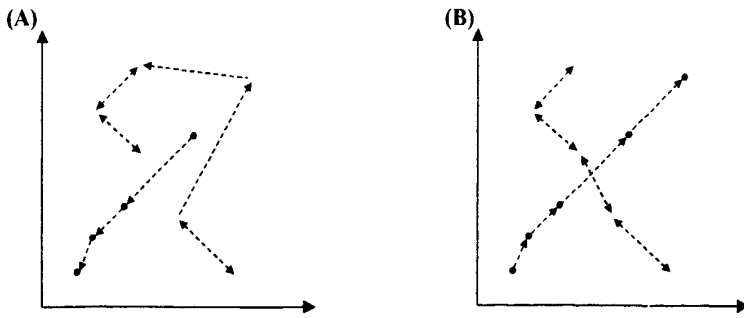


图2-7 时间信息如何用来分割物体的数集

图2-7中时间的转换用箭头表示，图(A)和(B)显示了两种不同的时间转换模式。分割数集需要识别通过时间转换连接在一起的点。两幅图中空心点表示一个物体，实心点表示另一个物体。

但是哺乳动物并不用此类方式解决这一问题。它们观察到的外部世界的图像是连续的，当它们和食物之间产生相对运动时在它们的视网膜上产生了不同的图像，然而这些图像在时间上是接近的。这意味着不同图像中的同一物体比不同图像中的不同物体更可能在邻接时间上连续发生。此外，因为世界中的物体运动遵循一定的物理定律，如惯性定律，两个连贯的图像更可能是来自同一个物体。这一信息已足够为分离数集消除模糊性。图2-7展示了时间信息如何帮助我们分离属于不同数集的点。

虽然不同的图像没有被标记，但是它们在时间上接近的事实可以作为它们是由同一物体产生的证明被用在学习的过程中，所以时间可以被看作是学习中的一种监督。

物体的不同图像属于同一数集，数集可以被看作是这一物体的不变表示。物体识别可以看作是输入图像确定其所属物体的数集。

一旦通过借助于时间信息的学习将这些数集分离，那么对于静止图像的识别也就变成了很简单的工作。只有使用无监督方式分离数集的过程需要时间信息，也就是说，在识别过程中时间信息只能起到帮助改进识别准确性的作用。

通过使用时间作为监督我们能够分离物体A和B的数集，但是对于物体A数集的学习并不能对物体B的数集的学习提供帮助。哺乳动物的视觉系统却很擅长借助于一物体识别的学习来学习另一物体。很多情况下我们能归纳一种我们从未见过的物体。因此实现视觉系统的这一能力比实现无监督学习和使用时间作为监督更加重要。

这就是要使用层级结构的原因。世界中的大部分物体共享着同样的组成模块，假设这个系统能够学会识别这些基础组成模块的不变性表示（数集），并且能学习又较小模块组成的较大模块的不变性表示，那么对不同物体的学习可以被认为是对不变性表示的基础模块的重新配置。在学习一个物体的同时我们也能够学习到其它物体的一些组成模块。而且组成模块的不变性识别对学习一个新物体有很大的帮助。

以上的论述可以被总结为两点：

- 使用时间信息来发现两个不同的图像（模式）是否属于同一个物体的数集，进

而完成不变性识别。

- 在层级结构中学习不变性识别这样在较高级别对较大物体进行不变性识别时可以利用已在较低层级习得的较小部分的不变性表示。

哺乳动物视觉系统中另一重要的方面是视觉的感知运动特性。动物的视觉与运动紧密联系，许多视觉问题只能在自发运动下才能解决。我们头脑中关于物体的模型不只是静止的图像，还包括由于运动产生的变化的图像。

贝叶斯网络是一种将贝叶斯概率方法和有向无环图的网络拓扑结构有机结合的表示模型，描述了数据库中数据项及其相互之间的依赖关系，贝叶斯网络就是根据各个变量之间的概率关系建立起来的图论模型，随着近年来数据库规模的不断扩大，研究发现应用贝叶斯方法进行数据库的知识发现要优于目前应用在数据库的知识发现的方法，这主要表现在以下几个方面。

1. 贝叶斯网络能够处理不完备数据集。这是传统的指导性学习方法所无法解决的问题，对于一般的指导性学习方法而言，必须知道所有可能的数据输入，如果缺少其中的某一输入就会对建立的模型产生偏差，贝叶斯方法则可以解决这个问题，因为贝叶斯网络反映的是整个数据域中数据间的概率关系，即使缺少某一数据变量仍然可以建立精确的模型

2. 贝叶斯网络是根据因果关系进行学习的。在数据分析处理中获得变量域的理解是十分重要的，而且贝叶斯网络可以在缺少插入值的情况下进行决策。

3. 贝叶斯网络和贝叶斯统计是紧密相关的。这促进了知识和数据域之间的关联关系，通常必须在知道处理数据域的先验知识的基础上才能建立正确的预测模型，由于贝叶斯网络具有语义的因果关系因而可以直接地进行因果先验知识的分析，所以在贝叶斯网络中可以获得较全面的先验知识。

4. 贝叶斯网络可以有效地避免数据溢出的情况。

2.4 贝叶斯网络

2.4.1 贝叶斯网络定义

贝叶斯网络亦称信念网络(Belief Network)，于是1985年由Judea Pearl首先提出。它是一种模拟人类推过程中因果关系的不确定性处理模型，其网络拓扑结构是一个有向无环图(DAG)。它的节点用随机变量或命题来标识，认为有直接关系的命题或变量则用弧来连接，表示结点间的依赖关系。例如，假设结点E直接影响到结点H，即 $E \rightarrow H$ ，则建立结点E到结点H的有向弧(E, H)，权值(即连接强度)用条件概率 $P(H/E)$ 来表示，该条件概率指明了变量与父结点之间的依赖关系。

2.4.2 贝叶斯网络图形表示

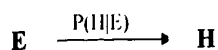


图 2-1 有两个结点的贝叶斯网络示意图

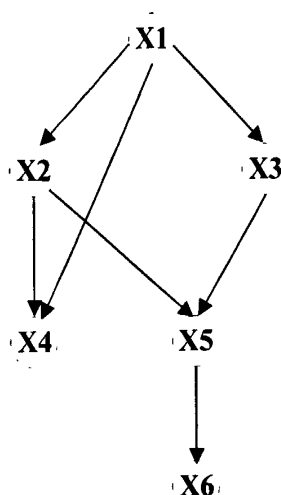


图 2-2 有 6 个节点的贝叶斯网络

2.4.3 贝叶斯公式和规则

先验概率和后验概率：

根据历史的资料或主观判断所得到的概率，该类概率没有经过检验，属于检验前的概率。分为客观先验概率和主观先验概率。

后验概率一般是指根据贝叶斯公式，结合调查等方式对先验概率进行修正后得到的比较符合实际情况的概率。

全概率公式：

设 A_1, A_2, \dots, A_n 是两两互斥的事件，且 $P(A_i) > 0$ ， $i = 1, 2, \dots, n$ ， $A_1 + A_2 + \dots + A_n = \Omega$ ，

另有一事件 $B = BA_1 + BA_2 + \dots + BA_n$ ，

$$P(B) = \sum_{i=1}^n P(A_i) \cdot P(B|A_i) \quad (2-1)$$

成为满足上述条件的 A_1, A_2, \dots, A_n 完备事件组

由此可以把全概率公式认为是由原因推结果，每个原因都对结果的发生有一定的作用，即结果发生的可能性与各种原因作用的大小有关系，全概率公式表示了它们之间的关系。

$$P(A_i|B) = P(A_i, B) / (P(A_1) \cdot P(B|A_1) + \dots + P(A_n) \cdot P(B|A_n)) \quad , \quad i=1,2,\dots,n \quad (2-2)$$

该公式于 1763 年由贝叶斯(Bayes)给出,它是在观察到事件 B 已发生的条件下,寻找导致 B 发生的每个原因的概率。

第三章 层级时间记忆模型

George. Dileep在记忆——预测结构的生物学基础上设计了满足这一结构的算法和数学模型^[7]。因为这种方法是一种使用了层级和时间记忆系统，所以被命名为层级时间记忆模型（Hierarchical Temporal Memory）简称HTM。这一方法可以被想象成现实世界的层级结构与记忆系统的映射。这种方法不只是把视觉问题看作是单纯的空间问题，而且使用时间维度在层级结构的每一个层次当中解决不变性表示的问题。

3.1 HTM 网络的拓扑结构

HTM网络具有树形拓扑结构，图3-1展示了一个三层的HTM网络。第一层有4个节点，第二层有2个节点，第三层有1个节点。输入信息首先被发送到第一层。

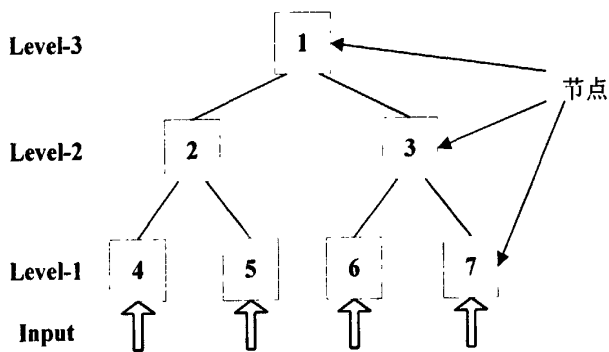


图3-1 一个含有7个节点，具有3层结构的简单HTM网络

输入信息首先被发送到处于网络层级中最底层的节点，接着在网络中被逐层向上传递直到顶层的节点。HTM网络中的节点是网络的基础算法与记忆模块。所有节点都有输入输出并且包含有相似的算法。节点也包含有记忆，这些记忆用来储存的节点环境信息。节点中的学习算法通过对输入空间的学习抽象出特征信息并储存在节点的记忆中。

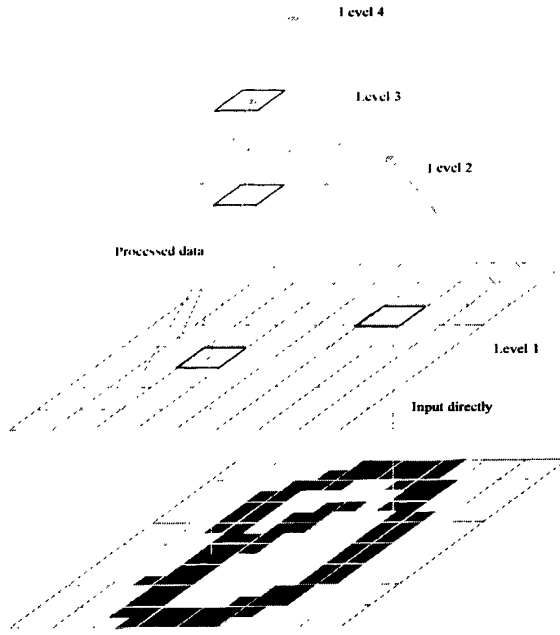


图3-2 用于进行不变性识别的简单网络结构

图3-2所示网络的level-1层有64个节点，按8x8排列。每个节点显示为一个小正方形。一个level-2节点的输入来自于4个level-1节点，4个level-2节点的输出进入一个level-3节点。

3.2 新大脑皮层层级结构与 HTM 网络层级结构的映射

这里将层级时间记忆的理论 and 运作方式与新大脑皮层不同层级的微通路的组织细节相对应。从皮层层级结构到HTM的层级结构、从皮质区域HTM的节点集合、从神经柱HTM节点中存储的模式都具有高度的对应。

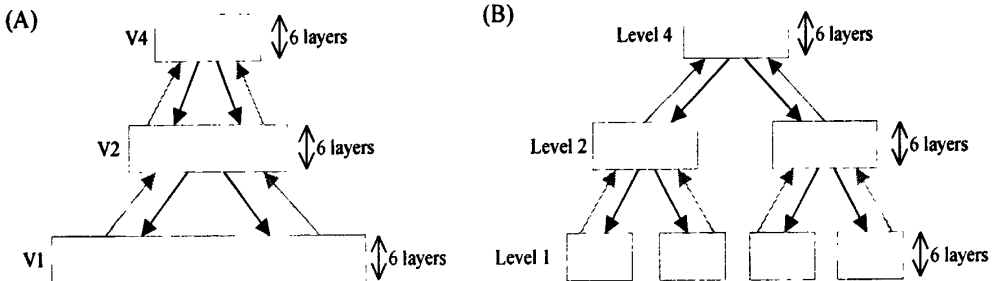


图3-3 信息在层级化的皮层区域间的传播

理想化的视皮层区域的层级结构连接应该如图3-3(A)所示，V1，V2和V4区是包含许

多神经柱的单个连续区域。但是显而易见的是较低的层级中一组神经柱的“激活”是独立其他的。例如在V1中左半部分可以看到一条垂直的线，右半部分可以看到一条水平的线。因为它们之间不会互相干扰，所以V1左侧的神经柱反映垂直线的同时右侧也可以反映水平线^[15]。

一个皮质区域可以被认为编码一组处于上下层级模式与序列关系中的模式和序列。这里的模式对应于HTM节点中存储的模式，序列对应于马尔可夫链。类似V1的区域能够编入许多互不排斥的模式。在理想模型中，这些模式集可以被划分为几个集合，每个集合内部的模式是互相排斥的，而属于不同集合的模式是互不排斥的。（互相排斥指一个假设的增强会使其他的假设强度减弱）

一组互相排斥的模式可以用一个随机变量来表示。在理想状态下，一个神经柱代表这个随机变量的特定值。在每个瞬间一个神经柱中的一组细胞的活动表现了一个特定模式“激活”的概率。与一组神经柱相连的前馈和反馈连接负责输送信念传播消息。皮质中任何位置的观测信息通过这些消息被传播到其它区域，并可以根据其他神经柱所包含的假设信息对概率值进行修改。

一个具有简化形式的HTM节点包含一组相互排斥的模式和马尔可夫链。皮质中同一区域数个模式同时激活的情况可以通过数个HTM节点实现。图3-3(B)显示了使用HTM实现的图3-3(A)中的皮质层级结构。在这一映射中V1区域应用了4个HTM节点，V2区域应用了2个HTM节点。通常情况下需要维持的非排斥模式的数量随着层级的上升而减少。所以高层的皮层区域在应用时需要较少的HTM节点。图4.1中只是简单示例，实际应用中要实现视皮层在V1和V2层各需要数千个HTM节点。

3.3 节点的运作方式

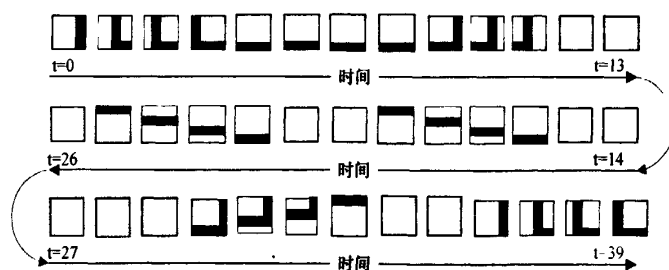


图3-4 输入序列

图中所示一个level-1节点接受的长输入序列中的一段。每个输入模式是一个包含16各元素的向量，代表4x4个像素的区域。

HTM网络使用物体的运动图像作为输入，训练时任一时刻的输入是训练影片中的一帧。一个level-1节点的视野范围是被输入图像中的一个4x4的像素阵列。通常物体会大于这一视野所以这一节点只能看到物体的一部分。当物体运动时，节点的视野范围内会

出现不同的模式，每个模式是一个16像素的向量。图3-4显示了一个level-1节点输入序列的一部分。

节点的运行包含两个阶段，在学习阶段节点观察输入它的模式并建立其在内部的表示，在这一过程中节点不产生任何输出。节点完整的学习过程是在感觉/推理阶段完成的。在这一阶段，节点为每个输入的模式产生输出，下面介绍节点在无噪声的状况下学习和感知阶段处理过程的细节。

3.3.1 节点的学习

学习过程中向节点的输入是一个长的模式序列。节点每次处理一个输入的模式。对于每个输入的模式，节点进行三个操作：

(1) 记忆输入的模式

节点内的存储部分保存视野范围内出现过的模式，并且赋予每一个模式一个不同的标签，如模式的编号。节点将由视野范围输入的模式与已经存储的模式加以对比，通过对比如果发现输入的是未存储的新模式的话，则把此模式加入到存储区中并为其添加一个标签。这一标签是任意的并不反映模式的任何属性。我们假设存储在节点中的模式是矩阵中的行，用 c 来表示这一矩阵。 c 中的每一行表示一个不同的模式，用 c_1, c_2, \dots 表示。

(2) 学习转换概率

这样节点构建了一个马尔可夫图（图3-5）。马尔可夫图中的顶点表示存储的模式，两个顶点之间的马尔可夫链表示的是两顶点之间发生转换事件的次数。当首次输入模式 i 与输入模式 j 被连续输入时，顶点 i 和 j 之间的链的转换事件次数被设置为1，此后每当 i 和 j 之间发生这样的转换时， i 和 j 之间的链就会被计数。

将每个链上的转换事件次数除以顶点上所有转换事件的总合，能够得到归一化的马尔可夫图。相应地，两个顶点之间的链就表示为两个模式间的转换概率。

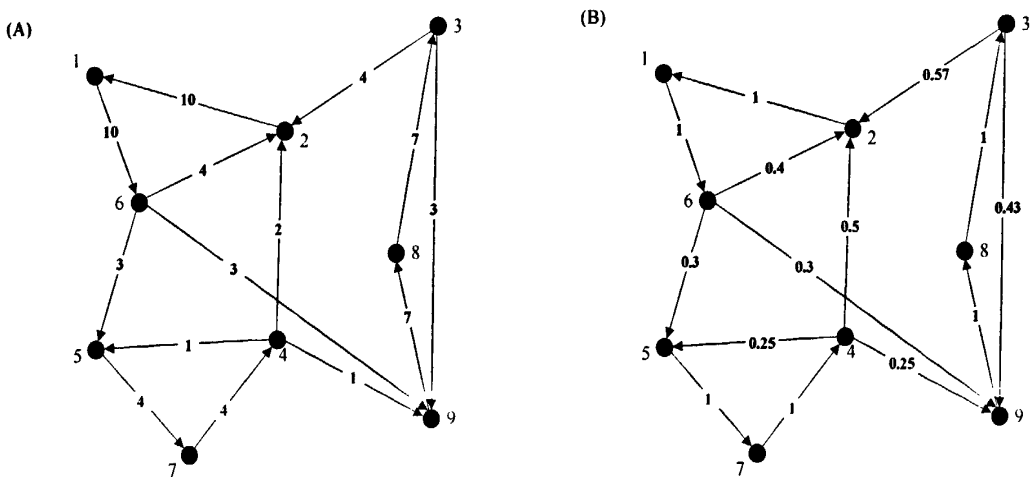


图3-5 通过训练得到的包含时间信息的马尔可夫图

图3-5 (A)所示一个未归一化的马尔可夫图。图中的顶点时节点中存储的模式对应的编号。一个顶点到另一个顶点之间的链标记的是由一顶点到另一顶点的转换事件次数。(B)是由(A)得到的归一化马尔可夫图。链上的标记表示了它链接的两个模式间的转换概率。

(3) 按时间分组

节点依据转换概率将马尔可夫图中的顶点集划分为一个时间组集，每个时间组是顶点集的子集。划分的目标在于把输入时间连续性较高的模式分到同一组中。

节点使用称为分层聚类法 (Agglomerative Hierarchical Clustering, AHC)^[49] 的分类方法形成时间组。AHC聚类算法将模式和模式的成对相似性作为输入，把单个的模式看作一个个的类，依据相似性进行合并，使类的数目逐渐减少，直到最后聚到所需的类数并且保证同一类中的模式较为相似。

两模式间的转换概率在这里被用来作为AHC算法中的相似性尺度。聚类是基于这一衡量模式连续发生可能性的相似性尺度进行的。时间上连续发生概率较高的模式会被分到同一聚类中，因此单个聚类也被称为时间组。

AHC初始时将每个储存着的模式作为一个单独的类，然后按照最大相似性的原则递归地合并聚类。最终将形成一个称为系统树图的树状结构。在树的顶部是包含所有模式的唯一聚类，底部是每一模式作为一单独聚类。按时间分组的目标分类位于树状系统顶部和底部之间的某一位置。这一位置是按一定适当的准则得到的，例如可以将得到的最大一个类的大小不超过某一特定值作为准则。节点有参数设置其分类（即时间分组）的最大数量，在系统树图中选择特定层次使得时间分组的数量最接近但小于设置的节点最大分组数则可以得到需要的目标分组。

这里使用一个简单的节点输入序列来阐述按时间分组的过程。图3-6所示的输入序列是由四个长度为3的子序列随机重复出现构成的。这样输入序列具有12个不同的模式。图3-6所示输入序列的一部分来阐释时间分组，在此长的输入序列在此例中随机选取4组长为3的子序列，记为“subseq”

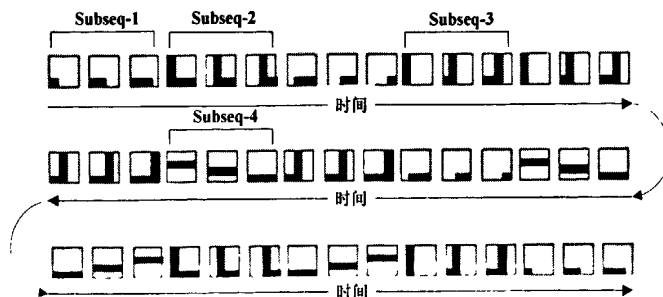


图3-6 输入序列

这些模式被储存到节点中并被赋予 c_1, c_2, \dots, c_n 作为标记。由这一序列得到的马尔可夫图如图3-7所示。与马尔可夫图对应的系统树图如图3-8所示。图3-9表示的是根据系统树图截取得到的时间分组。图3-10总结了节点在学习阶段的操作过程。

节点内的时间分组随着马尔可夫转换概率的更新而更新。节点的三个学习步骤——记录模式，形成归一化的马尔可夫图，按时间分组——是在节点的学习阶段实时进行着的。在时间分组足够稳定之后学习停止。

也可以以其他的方式完成这三部的学习过程，如可以首先记录所有模式，完成之后构建马尔可夫图，在马尔可夫图的概率充分稳定后进行归一化并完成按时间分组。

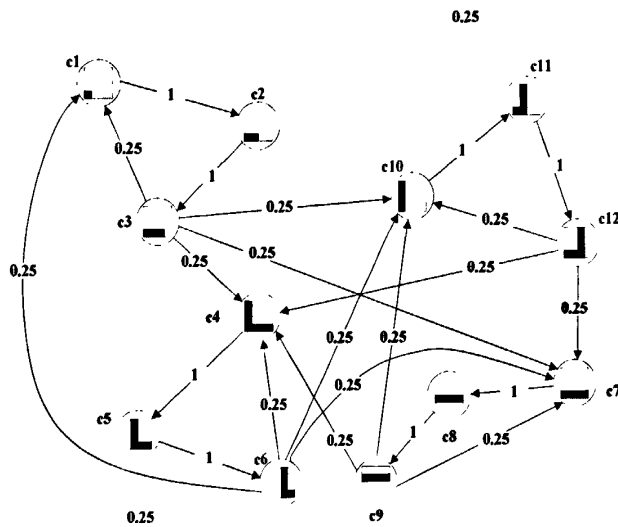


图3-7 由3-6序列训练得到的归一化的马尔可夫图

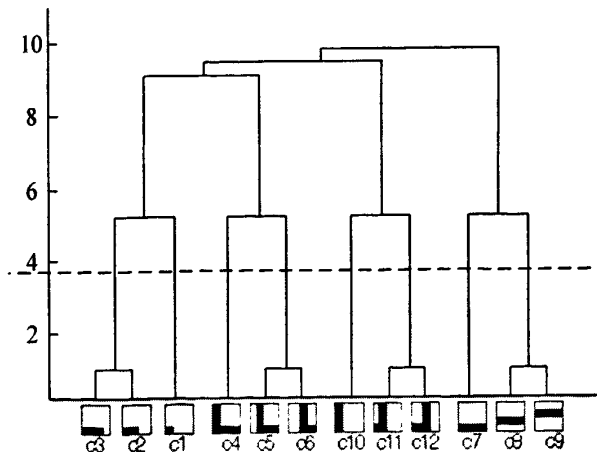


图3-8 显示了由图3-7中马尔可夫图得到的系统树图

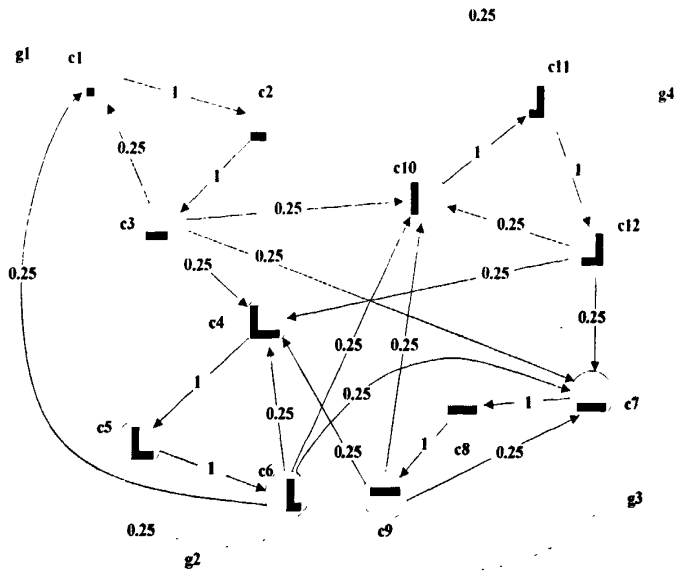


图3-9 图中所示根据系统树图水平线切割得到的时间分组

时间分组是通过把马尔可夫图划分成子图得到的。但在感知/推理阶段每个时间分组只被认为是一个模式的集合而不考虑分组内部的转换概率。因此本质上说，每个子图因舍弃了存储的序列信息而可以认为是零阶的马尔可夫链。

对于节点学习过程的总结

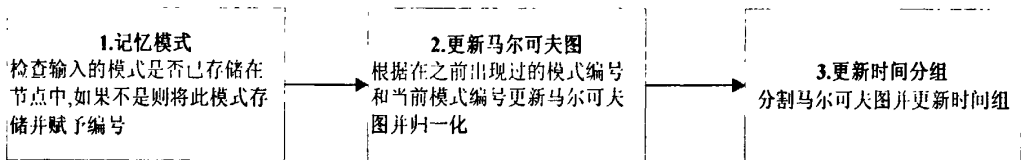


图3-10 对于节点的学习步骤

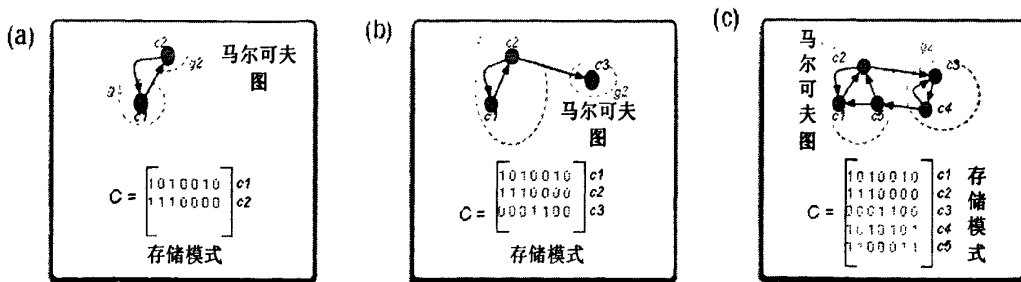


图3-11 描述了节点的学习过程

图3-11显示了节点的实时学习过程，输入节点的是长为7的模式，处于图(a)时节点已经存储了两个模式。这时的马尔可夫图有两个顶点，划分为时间分组 g_1 和 g_2 ；(b)中节点存储增加了一个模式，相应地，马尔可夫图增加了一个新的顶点，依然是两个时间

分组, 其中一组包含两个顶点; (c) 中节点结束了学习阶段, 存储了5个模式, 马尔可夫图有5个节点, 划分为两个分组。

3.3.2 节点的感知和推理

节点在完成学习阶段后进入感知/推理阶段。感知时的输入特征已经在学习阶段被节点记录, 并且感知时物体是运动的, 节点在其自身的视野范围内能够观察到物体的一部分。最终得到的模式是节点的输入。

在感知/推理阶段节点为每个输入模式产生一个输出。因为此阶段使用已习得的时间分组而并不会用到马尔可夫图, 所以节点既可以对连续的输入序列产生输出, 也可以对瞬时的独立输入产生输出。

对每个输入的模式, 节点产生一个向量, 向量中的每个分量表示此输入模式属于对应时间分组的程度。由于基本上输入的模式不会与节点中存储的某个模式完全相同, 因此需要求出输入模式与节点中已存储的每个模式的相似性。用 d_i 表示节点中存储的 i_n 模式与输入模式间的距离, 距离越远, 相似性越低。可以认为输入模式符合某一存储模式的概率符合高斯分布, 所以, 输入模式符合已存储的 i_n 模式的概率可使用 $e^{-d_i^2/\sigma^2}$ 计算, 这里 σ 是节点的一个参数。对每个存储的模式进行这样的计算能够得到输入模式与马尔可夫图中各个顶点的相似性。

输入模式属于各时间分组的程度是由它与组中各顶点最大相似性决定的。最终得到长度为时间分组个数的向量, 向量的每个分量表示此输入模式属于对应时间分组的程度。向量通过归一化使得各分量之和为 1, 这样输入模式属于各时间分组的程度也可以理解为相应的概率。

3.3.3 层次结构中节点的操作

网络结构中的节点是依层次进行训练的。首先对 level-1 的节点进行训练, 训练时并不产生输出。这时 level-2 以及更高层的节点处于关闭状态。一旦 level-1 的节点结束训练就会切换到感知/推理模式, 这时节点产生输出, 而这些输出即成为 level-2 节点训练的输入, 对 level-2 的节点进行训练。网络就是这样逐层向上进行训练的。

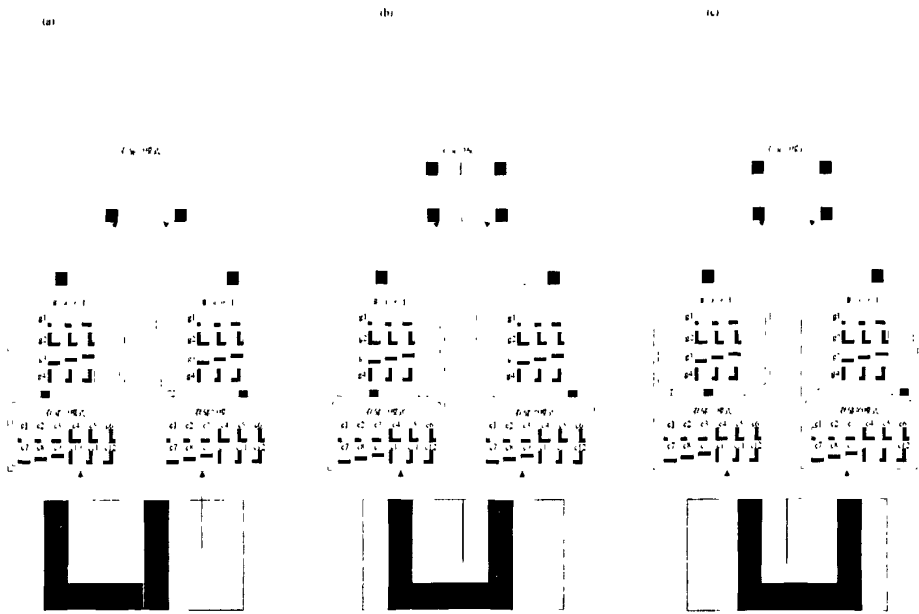


图3-13 层次结构中节点的运作方式^[46]

这里使用如图3-13 (a) 这样的一个两层简单网络为例子介绍层次结构中节点的操作。图中level-1有两个节点，level-2有一个节点。网络的输入是一个4 x 8像素的图像。Level-1的每个节点对应于图像的一半，即一个4 x 4像素的模式。两个level-1节点的内部结构是相同的，都包含12个模式和4个时间分组。图中level-1节点处于感知/推理阶段，level-2节点处于学习阶段。

处于第一层的节点，节点的输入直接来自网络的输入。较高级别的节点的输入其直接相连的低一级的数个节点输出的串联。图3-13中的每个level-1节点具有长度为4的输出，所以level-2节点的输入为将两个子节点输出串联起来的长度为8的模式。

为简化表述这里假定输入网络的模式与level-1中存储的某个模式能够完全吻合，也就是对于一个输入模式，它属于包含与他相同的模式的时间分组的概率为1，而属于其他时间分组的概率为0。因此，在图3-13中level-2的节点的输入模式中应包含两个‘1’，其余位为‘0’。

图3-13显示了一个两层的示例网络在 $t=0$, $t=1$, $t=2$ 时刻的输入, 通过将两个level-1节点的视野范围相结合可以看出输入的是一个移动的‘U’形图像的模式序列。图中可以看出右下节点在这三个时刻的输出同为 $[0, 0, 0, 1]$ ，输出并没有随输入的变化发生改变, 因为输入的模式 c_{10}, c_{11}, c_{12} 属于右level-1节点的同一个时间分组 g_4 。同样, 左level-1节点在这三个时刻的输出也同为 $[0, 1, 0, 0]$ 。

在任何时刻，level-2节点的输入是由该节点的子节点的输出串联组成的。在上述的三个时刻，level-2节点的输入为 $[0, 1, 0, 0, 0, 0, 0, 1]$ ，是由左子节点输出的 $[0, 1, 0, 0]$ 和右子节点输出的 $[0, 0, 0, 1]$ 组成的。由于网络接触到了三幅不同的U的图像，但是到了level-2节点的输入却是相同的，这一输入即是‘U’三个变换位移的

不变性表示。

Level-2节点记忆模式的过程与level-1节点是相同的。在学习过程中如果节点发现一个未被存储的新模式则将其添加到节点的存储中并赋予其一个编号。在从 $t=0$ 到 $t=2$ 时刻的输入序列中，一个新的模式被添加到了level-2的存储中。

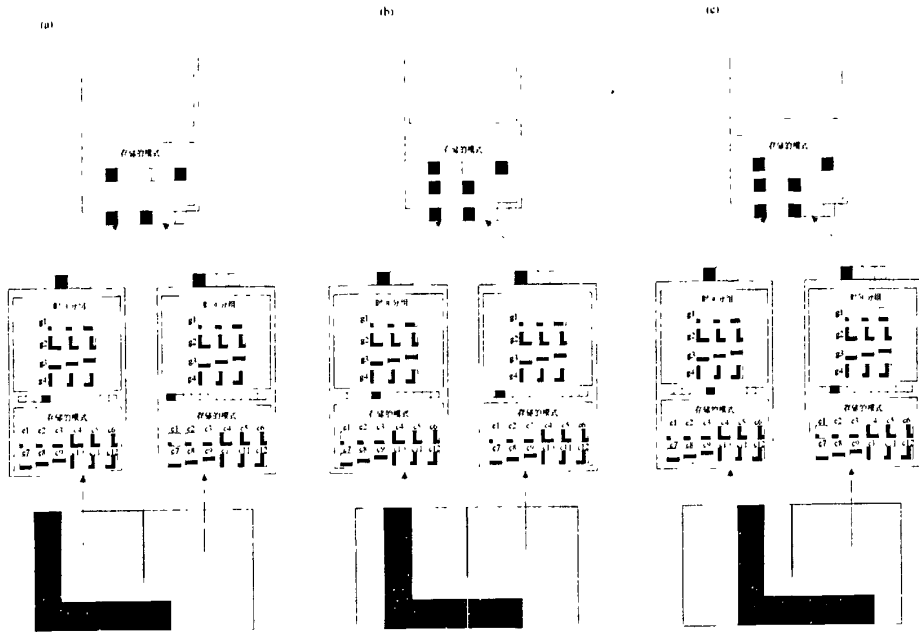


图3-14 输入序列变为了向右移动的‘L’，新的输入序列被加入level-2节点之中^[46]

图3-14显示了另一个模式的序列，其对应于由 $t=10$ 到 $t=12$ 时间段的输入和输出。在这一时间段中图形‘L’在由level-1节点组成的视野内移动。

如同‘U’模式的情况一样，‘L’模式在自身变化的情况下向level-2的输入是相同的，这一输入也是‘L’在这一时间段内变化的不变性表示。但是‘L’序列向level-2节点的输入与‘U’序列向level-2节点的输入是不同的。也就是说，level-2的节点在能够区别‘U’和‘L’不同模式的同时保证它们模式自身的不变性识别。

对于学习转换概率和按时间分组的方法，level-1节点与更高层的节点是相同的。在多层网络的学习过程中这一过程按层级重复地进行。

3.4 噪声处理

3.4.1 有噪声输入的学习

前面所述的存储过程适用于无噪声的情况。Level-1的节点与更高层次节点为了处理噪声所作的修改有所不同，这里先介绍level-1的修改，然后介绍更高层次的修改。

当输入具有噪声时，就需要在聚类之前加入一步预处理过程。预处理的算法是基于

输入的模式属于一定的向量空间。向量空间中的不同模式使用欧几里得距离来判断属于哪个模式的集群。这样欧几里得距离彼此接近的模式会被划如同一个集群当中。

节点使用 k 均值聚类算法使输入的模式聚集成 k 个集群, k 在这里被设计为节点的一个参数。节点通过 k 均值算法从输入模式中得 k 个集群, 这些集群中心将会存储到节点当中并且赋予不同的标签标记。将 k 个集群中心代替前面所述无噪声学习时的存储的模式, 之后的学习步骤与无噪声学习时相同。

向更高层节点的输入是反映了网络输入的模糊性的概率分布。在学习过程中, 高层节点通过稀疏化来处理这种模糊性, 这种稀疏化是通过为每个子节点选取一个“优胜者”实现的。子节点向上输入的数据中, “优胜者”的位置设为1, 其他位置设为0。通过这一预处理, 在节点进行学习时可以使用与无噪声学习时同样的方法。

3.4.2 有噪声输入的感知/推理

在之前4.3.2介绍level-1节点感知/推理过程中, 已经假设输入的模式不是完全符合已存储的模式, 所以level-1节点可以直接处理有噪声的情况, 其输出能够反映输入模式的噪声情况。这时输出是反映有噪声输入属于每个事件分组可能性的概率分布。

高层节点计算输出的方法与level-1节点有轻微的不同, 如图3-13中所示向level-2节点输入的向量, 来自左子节点的‘1’对应它内部的group-2, 来自右子节点的‘1’对应它内部的group-4。如果来自子节点的输入是概率分布的情况, 那么来自两节点输出串联的这一个输入的模式的发生概率就与左子节点输出向量的第2个分量和右子节点的第4个分量之的乘积成正比。

以公式表示, 向level-2输入的第 i^{th} 个模式表示为 $c_i = [m'_1, \dots, m'_N]$, 其中 m'_k 表示向量中非零的位置, N_c 表示子节点的个数。第 i^{th} 个模式发生的概率为

$$P(c_i) = \gamma \prod_{k=1}^{N_c} \text{input}(m'_k) \quad (3-1)$$

其中 γ 是归一化常数。

当得到所有已存储模式的概率分布之后, 其余的计算步骤与level-1的推断步骤相同。从已存储模式的概率分布, 特定时间分组内概率最高的模式的概率将作为该分组的发生概率。将所有时间分组的概率归一化后得到时间分组的概率分布。

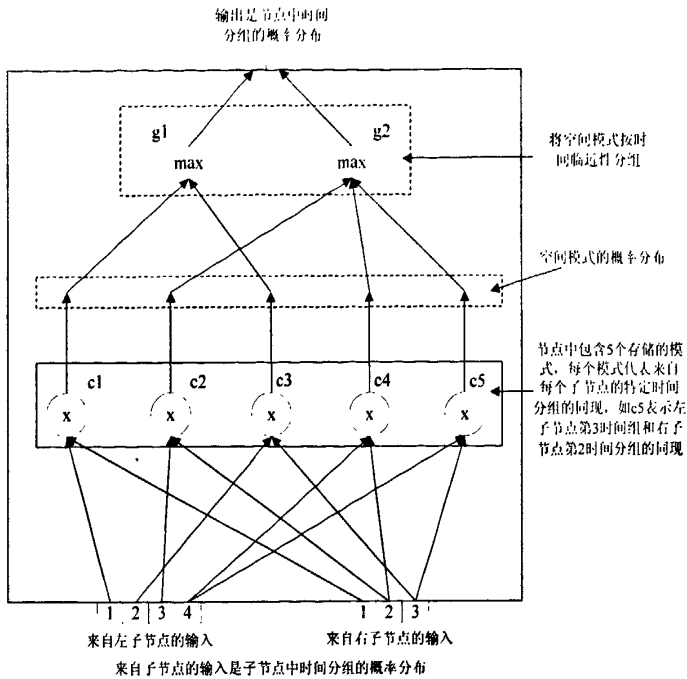


图3-15 说明了高层节点的结构和一般推理机制

每个模式可以认为能够用一个类似神经元的单元表示。神经元与输入的连接类似于模式与输入的连接，如 c_2 与左子节点输入消息的3rd位置有连接，同时连接着右子节点的2nd位置。这意味着 c_2 代表了左子节点第3组和右子节点第2组的同现。同现的神经元通过对输入进行相乘得到输出，类似地，每个时间组可以被认为是一个神经元。时间分组由它连接到的模式决定。在图中时间组 g_1 由模式 c_1 和 c_3 组成，时间组 g_2 由模式 c_2 、 c_4 和 c_5 组成。时间组的输出通过在输入中求最大值得到。

这种不确定性的传播在识别时比“赢家通吃” (winner-take-all) 规则更具优势。模糊性能在更高层次被解决因为只有一部分的低层时间分组式高层的有效模式输入。贝叶斯信念传播 (Bayesian belief propagation) ^[48] 在网络中传播这种不确定性以获得最优识别性能。同时沿着网络层级向下传播证据消除下层的的不确定性。

3.5 HTM 网络中的信念传播

贝叶斯信念传播首先由Judea Pearl提出并在他所著《Probabilistic Reasoning in Intelligent Systems》一书中作了详细的论述^[48]。HTM中的信念传播公式是通过贝叶斯信念传播的公式进行改进得到的。HTM像贝叶斯网络一样将随机变量间的关系编入自身。HTM中这些随机变量对应于网络各个层次的模式和马尔可夫链。

向网络中的输入通常叫做证据。例如向网络呈现的一个新的图像就叫做网络的证据。网络在层级中传播这些证据来调整每个节点的信念状态。信念状态被定义为每个节

点中的模式的后验概率。

信念传播算法存在两个主要变体，它们基于同样的模型解决不同的要求。一个变体叫做sum-prop，利用证据计算每个节点中的信念状态；另一个叫做max-prop，也称为信念修正，这一算法利用证据寻找节点状态的分配的最佳可能组合。信念修正计算的结果被认为是当前证据的最可能解释^[18]。这两种变体具有细微的不同，sum-prop的方法，节点中的信念状态计算的是不假设其他节点处于特定状态情况下节点中信念状态。max-prop的方法，节点的信念状态需要处于最佳配置下，这就要假设其他节点处于特定的状态。

一个完成学习的HTM节点可以认为是存储了一组模式和一组关于模式的序列（马尔可夫链）。例如对应于V1中不同区域部分的HTM节点中不同的模式能够代表不同的垂直方向边缘和不同的水平方向边缘。马尔可夫链中的一条可能能够表现垂直边缘的运动序列而另一个可能表现水平边缘的运动序列。

网络节点中的信念状态反映了这一证据。当呈现一幅新图像时，这些信息通过信念传播沿着层级向上和向下传递，节点中的信念状态发生改变来匹配新证据。

总体来说一个HTM节点获得的来自于子节点的消息表现了该子节点对于其所含马尔可夫链的确定程度，而这个节点又将这一消息转换为对其自身所含模式的确定程度。基于收到的消息的历史，节点可以计算其所含马尔可夫链的确定程度。之后这一信息被传送到更高层节点。节点从父节点接受的是父节点对该节点各马尔可夫链的确定程度。随着信息自顶向下的传递各模式的概率分布被逐渐确定，这一概率分布将参与到关于节点自身模式和马尔可夫链确定程度的计算当中。静态情况的信念传播公式

符号

C^k k^{th} 节点中代表模式的随机变量，在某些情况下也表示 k^{th} 节点中代表模式的随机变量的集合。

C_i^k k^{th} 节点中的 i^{th} 模式

G^k k^{th} 节点中代表马尔可夫链/时间分组的随机变量，在某些情况下也表示 k^{th} 节点中代表时间分组的随机变量的集合。

g_i^k k^{th} 节点中的 i^{th} 马尔可夫链/时间分组

e^- 和 \bar{e} 来自节点下面的证据，由所有这一节点的后代节点的所有观察证据组成。

e^+ 和 \bar{e}^+ 来自节点上面的证据，由所有这一节点的前辈节点的所有观察证据组成。

$P(e^- | G^k)$ 在节点 k 的马尔可夫链/时间分组给定条件下来自下面的证据的条件概率，这是一个长度等于时间分组数的向量。

$P(e^- | g_i^k)$ 在节点 k 的 i^{th} 时间分组给定条件下来自下面的证据的条件概率，这是一个标量。

$P(e^- | C^k)$ 在节点 k 的模式给定条件下来自下面的证据的条件概率，这是一个长度等于模式数的向量。

$P(e^- | c_i^k)$ 在节点 k 的 i^{th} 模式给定条件下来自下面的证据的条件概率，这是一个标量。

$P(C^k | G^k)$ 矩阵的行对应于马尔可夫链/时间分组，列对应于存储的模式。矩阵的 $(i, j)^{\text{th}}$ 元 $P(c_i | g_j)$ 表示给定马尔可夫链/时间分组 g_j 条件下，观察到模式 c_i 的条件概率。

λ^k 节点 k^{th} 向父节点发送的消息，是一个与 $P(e^- | G^k)$ 成正比的向量。 $\lambda^k(r)$ 是该向量的 r^{th} 分量，正比于 $P(e^- | g_r^k)$

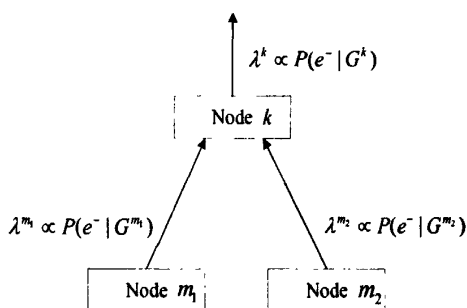


图3-16 HTM网络的一部分

完成学习的节点

存储的模式和马尔可夫链/时间分组构成了完成学习的节点。

模式 c_i 子节点的马尔可夫链/时间分组的特定同现，可以用向量 $[r_1^{m_1} \dots r_{i^{m_m}}]$ 表示， r_i 表示子节点的时间分组的编号。例如 $M = 2$ ，来自子节点 m_1 的第2组和子节点 m_2 的第5组同现组成 c_i ，那么 c_i 代表的模式就为 $[2, 5]$ 。

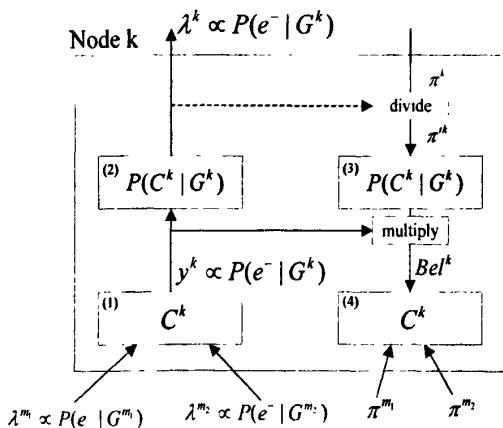


图3-17 静态情况下HTM节点的信念传播

对于静态情况的信念传播，需要一个由时间分组得到的条件概率矩阵，用 $P(C|G)$ 表

示。矩阵的 i^{th} 行对应节点中的 i^{th} 时间分组， j^{th} 列对应 j^{th} 模式。 i^{th} 行使用如下方式计算：属于这组的模式的对应的位置对该模式的发生次数进行累加，其他位置累加0，之后对该行归一化得到 i^{th} 行。 C 和 $P(C|G)$ 共同组成了完成学习的HTM节点。

信念传播计算

图3-16展示了HTM网络的一部分，这部分包含有3个节点，节点 k 有两个子节点—— m_1 和 m_2 。考虑一般性假设节点 k 具有 M 个子节点，以 m_1, m_2, \dots, m_M 表示。

此节点从子节点中获得 M 个消息。来自子节点 m_i 的消息 λ^m ：

$$\lambda^m \propto P(e^- | G^m) \tag{3-2}$$

这一消息正比于给定节点 m_i 的时间分组 G^m 的条件下来自下面的证据的条件概率。这是一个长度等于节点 m_i 时间分组数的向量。

前馈计算

获得这些消息后，首先计算模式的确定性程度，用 y 表示且与 $P(e^- | C)$ 成正比。向量 y 的 i^{th} 分量对应于模式 c_i 。一般情况下可以认为模式具有 M 个分量，以向量 $[r_1^m, \dots, r_M^m]$ 表示， r 表示对应子节点的时间组编号。 y 的 i^{th} 分量用以下方法计算

$$y_i(i) = \alpha_i \prod_{j=1}^M \lambda_j^m (r_j^m) \tag{3-3}$$

α_i 任意度量常数。这个度量常数是设定用来防止消息产生浮点下溢的。

这里假设来自节点的证据在给定模式的条件下是相互独立的。因为 $P(e^- | c_i) = \prod_{j=1}^M P(e^- | g_j^m)$ 且 λ^m 对于所有 i 与 $P(e^- | G^m)$ 成正比，所以这一计算能保证 y 与 $P(e^- | C)$ 成正比。

节点计算后输出 y ，正比于 $P(e^- | G)$ ，向量 y 的长度为 N_g ，等于节点中时间分组的数量。向量的 i^{th} 分量计算如下

$$\lambda^k(i) = \sum_{j=1}^N P(c_i | g_j) y(j) \tag{3-4}$$

因 $P(e^- | g_i) = \sum_{j=1}^N P(c_j | g_i) P(e^- | c_j)$ ，所以 λ^i 与 $P(e^- | G_i)$ 成正比。

反馈计算

反馈计算使用来自父节点的自上而下的消息 λ^k 计算要发送到子节点的消息。这个过

程中节点的信度是通过将模式的前馈概率与模式的反馈概率相乘得到的。首先用如下方式计算 π'

$$\pi^{k'}(i) = \pi^k(i) / \lambda^k(i) \tag{3-5}$$

自上而下的模式概率为

$$z(i) = \sum_{g, g \in G} P(c_i | g) \pi'(j) \tag{3-6}$$

节点内模式的信度是自上而下和自下而上的概率的乘积

$$Bel(c_i) = y(i) \times z(i)$$

向子节点传播的 π 消息通过对信念分布和模式的计算得到

$$\pi^{child}(g_m^{child}) = \sum_{c_i \in C} I(c_i) Bel(c_i) \tag{3-7}$$

其中

$$I(c_i) = \begin{cases} 1, & \text{if } g_m^{child} \text{ is a component of } c_i \\ 0, & \text{otherwise} \end{cases} \tag{3-8}$$

信念修正的计算可以通过将公式A. 3, A. 5 and A. 7中的求和替换为求最大值得到。

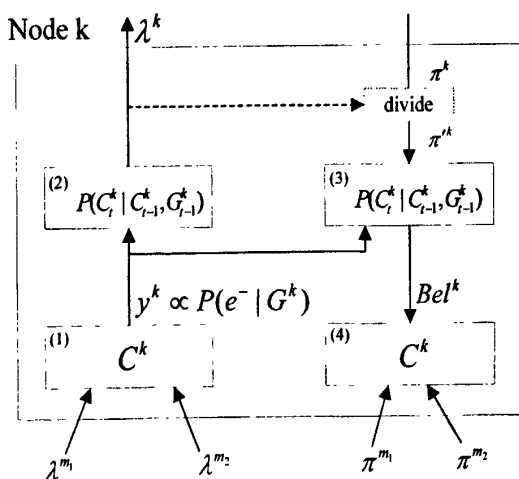


图3-18 动态情况下HTM节点的信念传播

3.5.1 动态情况下的信念传播公式

当证据以序列化的方式呈现时，就需要使用动态规划的方法[5, 53, 93]得到顺序推理的公式。

符号

图A. 2显示了动态情况下节点进行信念传播计算的框图。

$e_0^-, \bar{e}_0^-, \bar{e}_0^+$ 这些代表从0到 t 时刻相关变量的可能的序列组合

$\alpha, Bel, \gamma, \beta$ 所有的内部变量在每个时步都要进行更新，特定的时刻用下标 t

表示

$c(t)$ 代表 t 时刻模式所有可能得到的值，马尔可夫链也有类似的符号

$c_i(t)$ 代表 t 时刻的模式 c_i

$P(C_t^i | C_{t-1}^i, G_{t-1}^i)$ 代表节点中马尔可夫链集的所有内部马尔可夫链转换概率

模式和马尔可夫链

动态情况下的模式与静态情况下相同。静态情况下的时间分组是零阶马尔可夫链，而动态情况下是一阶马尔可夫链。马尔可夫链集的转换概率是已习得的数据，用 $P(C_t^i | C_{t-1}^i, G_{t-1}^i)$ 表示。

动态规划公式

节点 k 随时间接收到消息，为简化计算假设自底向上的消息到达节点的时间符合时步 $t=0, t=1, \dots$ 来自父节点的消息到达的时间间隔是 τ_k 的倍数，这里 τ_k 是节点 k 的时间常数。类似地，假设自底向上的消息是以 τ_k 为时间跨度传递。

假设自顶向下的消息在 $t=0$ 时刻到达，而来自子节点的自下而上的消息同时到达。我们得到从 $t=0$ 到 τ_k 时刻用于进行节点内部状态和输出计算的更新公式。

$$\begin{aligned}
 Bel_t(c_i) &= P(c_i(t) | \bar{e}_0^-, \bar{e}_0^+) \\
 &= (1 / P(\bar{e}_0^-, \bar{e}_0^+)) \sum_{g_r \in G^t} \sum_{c_0^i \in C_0^i} P(\bar{e}_0^+ | c_0^i, g_r, \bar{e}_0^-) P(c_0^i, g_r | \bar{e}_0^-) \\
 &= (1 / P(\bar{e}_0^-, \bar{e}_0^+)) \sum_{g_r \in G^t} P(g_r | \bar{e}_0^-) \sum_{c_0^i \in C_0^i} P(\bar{e}_0^+ | c_0^i, g_r, \bar{e}_0^-) P(c_0^i, g_r | \bar{e}_0^-) \\
 &\propto \sum_{g_r \in G^t} P(g_r | \bar{e}_0^-) \beta_t(c_i, g_r) \tag{3-9}
 \end{aligned}$$

动态规划变量 β 定义为

$$\beta_t(c_i, g_r) = \sum_{c_0^i \in C_0^i} P(\bar{e}_0^+ | c_0^i, g_r, \bar{e}_0^-) P(c_0^i, g_r | \bar{e}_0^-) \tag{3-10}$$

所以，更新公式变成

$$\beta_i(c_i, g_r) = P(\bar{e}_i | c_i(t)) \sum_{c_j(t-1) \in C^t} P(c_j(t) | c_j^{t-1}, g_r) \beta_{i-1}(c_j, g_r) \quad (3-11)$$

初始状态 $\beta_0(c_i, g_r) = P(\bar{e}_0 | c_i(t=0)) \sum_{c_j(t-1) \in C^t} P(c_j(t=0) | g_r, \bar{e}_0)$ 可以认为是马尔可夫链的初始分布。

自底向上传播的消息计算如下

$$\begin{aligned} P(\bar{e}_0(t) | g_r(t)) &= \sum_{c'_0} P(\bar{e}'_0, c'_0 | g_r) \\ &= \sum_{c'_0} P(\bar{e}'_0 | c'_0) P(c'_0 | g_r) \\ &= \sum_{c'_0} P(\bar{e}'_0 | c'_0) P(\bar{e}_i | c_i) P(c'_0, c_i | g_r) \\ &= \sum_{c'_0} P(\bar{e}_i | c_i) P(c_i | c_{i-1}, g_r) P(\bar{e}'_0 | c'_0) P(c'_0 | g_r) \\ &= \sum_{c_i \in C^t} P(\bar{e}_i | c_i(t)) \sum_{c_j \in C^t} P(c_j(t) | c_j(t-1), g_r) \sum_{c'_0} P(\bar{e}'_0 | c'_0) P(c'_0 | g_r) \\ &= \sum_{c_i \in C^t} P(\bar{e}_i | c_i(t)) \sum_{c_j \in C^t} P(c_j(t) | c_j(t-1), g_r) \alpha_{i-1}(c_j, g_r) \quad (3-12) \end{aligned}$$

其中 α 是动态规划变量，更新公式为

$$\alpha_i(c_i, g_r) = P(\bar{e}_i | c_i(t)) \sum_{c_j(t-1) \in C^t} P(c_j(t) | c_j(t-1), g_r) \alpha_{i-1}(c_j, g_r) \quad (3-13)$$

自底向上输出的消息计算方法是

$$\lambda(g_r) = P(\bar{e}'_0 | g_r(t)) \alpha \sum_{c_j(t) \in C^t} \alpha_i(c_j, g_r) \quad (3-14)$$

3.5.2 HTM 网络中的反馈传播

识别是通过将图像作为网络的输入，然后在层级中使用贝叶斯信念传播向上传播证据。对于任何呈现过的图像，信息也会在网络中自上而上传播。这是通过信念传播公式实现的，反馈传播用来消除图像中的噪声或驱动注意力。下图面熟了网络中反馈传播的过程

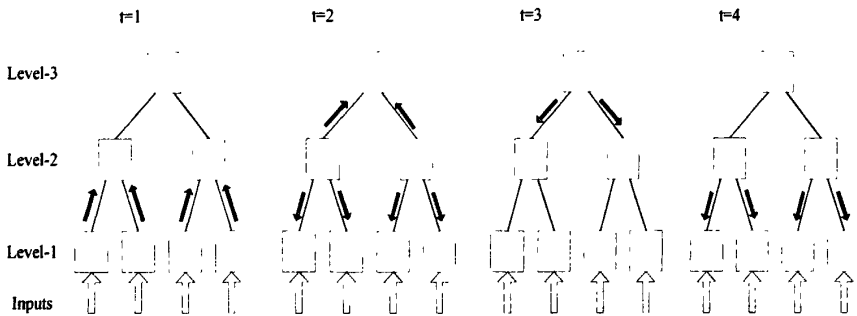


图3-19 网络中的信念传播

信念传播消息可以认为是网络中传播的扰动。在时刻 $t = 0$ 一副新图像被输入网络，这对 level-1 节点的均衡产生了扰动并向上传播；在 $t = 1$ 时刻，消息到达 level-2 节点，触发了反馈消息的传播；在 $t = 2$ 时刻，前馈消息到达顶层节点，同时从 level-2 出发的反馈消息到达 level-1 节点；在时刻 $t = 3$ ，来自 level-3 节点的反馈消息到达 level-2 节点，触发另一波的反馈传播；在 $t = 4$ 时刻，结合了整个网络的证据的反馈到达 level-1 节点。

3.5.3 HTM 节点中的信念传播公式

表3-1 HTM节点内部信念传播的计算过程

1) 计算模式 概率	$y_i(t) = P(\tilde{e}_i c_i(t)) \propto \prod_{j=1}^M \lambda_j^m(r_j^m) \quad (3-15)$ 这里模式 c_i 是指来自子节点1的 r_1^m ，来自子节点2的 r_2^m ，……来自子节点M的 r_M^m 同时发生。
2) 使用动态 规划马尔可 夫链的前馈 概率	$\lambda_i^k(g_r) = P(\tilde{e}'_i g_r(t)) \propto \sum_{c_j(t) \in C^k} \alpha_i(c_j, g_r) \quad (3-16)$ $\alpha_i(c_j, g_r) = P(\tilde{e}'_i c_j(t)) \sum_{c_j(t-1) \in C^k} P(c_j(t) c_j(t-1), g_r) \alpha_{i-1}(c_j, g_r) \quad (3-17)$ $\alpha_0(c_j, g_r) = P(\tilde{e}_0 c_j(t=0)) P(c_j(t=0) g_r) \quad (3-18)$
3) 计算模式 的信念分布	$Bel_i(c_i) \propto \sum_{g_r \in G^k} P(g_r \tilde{e}_0) \beta_i(c_i, g_r) \quad (3-19)$ $\beta_i(c_i, g_r) = P(\tilde{e}'_i c_i(t)) \sum_{c_j(t-1) \in C^k} P(c_i(t) c_j(t-1), g_r) \beta_{i-1}(c_j, g_r) \quad (3-20)$ $\beta_0(c_i, g_r) = P(\tilde{e}_0 c_i(t=0)) P(c_i(t=0) g_r, \tilde{e}_0) \quad (3-21)$

4) 计算要发送到子节点的消息	$\pi^{child}(g_m) \propto \sum_{c_i} I(c_i) Bel(c_i) \quad (3-22)$
	$\text{其中 } I(c_i) = \begin{cases} 1, & \text{if } g_m^{child} \text{ is a component of } c_i \\ 0, & \text{otherwise} \end{cases} \quad (3-23)$

图3-20详细介绍了公式中的符号和公式的意义。

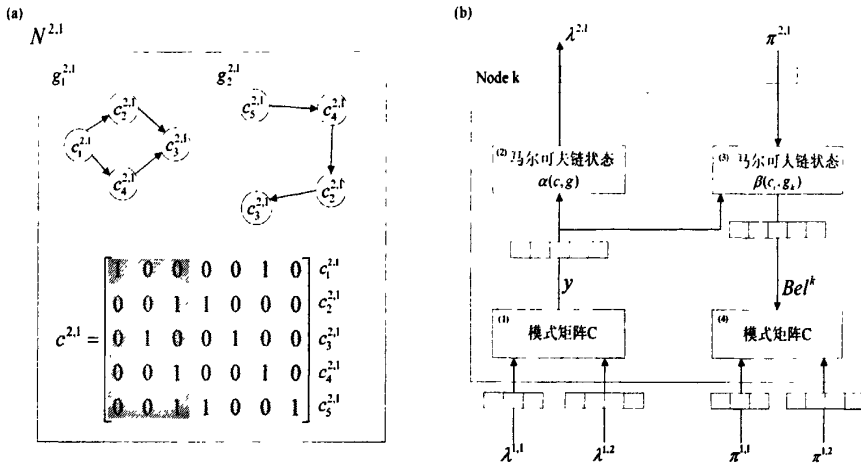


图3-20 节点结构和不同层级节点间的信息传播

图3-20 (a) 一个经过学习，具有5个模式和2个马尔可夫链的推理节点的结构。假设这个节点处于level-2并且标记为 $N^{2,1}$ 。这一节点具有2个子节点，子节点1有3个马尔可夫链，子节点2有4个马尔可夫链。每个存储的模式反映了一种自节点的马尔可夫链的同现。模式中来自不同子节点的部分用不同背景色加以区分。(b) 节点内的矩形对应公式表4.1的4个处理过程。图中的结点有两个自底向上的输入消息，对应于两个子节点；有两个自上而下的输出发送到两个子节点。

在公式中模式使用 c_i 表示，马尔可夫链使用 g_j 表示。图3-20(a) 中的HTM节点包含5个模式和2个马尔可夫链。马尔可夫链 g_j 的转换概率矩阵是通过公式 $P(c_i(t)|c_i(t-1), g_j)$ 表示的。节点中的每个模式表现了其子节点中时间分组的同时发生，这也由公式4.1和4.8体现。

节点从子节点接收前馈消息，随后将前馈消息传播到它的父节点。前馈传递消息用 $\lambda^{source\ node\ index}$ 表示。同样地，节点从父节点接收反馈消息再将反馈传播消息传递到它的子节点。反馈消息用 $\pi^{destination\ node\ index}$ 表示。表4.1中的公式描述了怎样通过输入消息得到输出消息。从节点的角度，前馈消息携带了来自底层的证据的信息。在任何时间点，来自底层的证据用 e_i^- 表示，类似的来自父节点的消息用 e_i^+ 表示。

公式3-15描述了怎样通过来自子节点的消息计算其模式的可能性。在 t 时刻模式 c_i 的可能性由公式 $y_i(t) = P(e_i^- | c_i(t))$ 表示。各个模式的可能性通过计算消息中与该模式对应部分的乘积得到。

公式3-16中, t 时刻马尔可夫链 g_t 的可能性用 $P(\tau_{e_0} | g_t(t))$ 表示, τ_{e_0} 表示从时刻0到时刻 t 自底向上的证据序列。这反映了马尔可夫链的可能性依赖于节点接收到的输入序列。公式3-17和3-20中定义的变量 α 和 β 是随着时间变化不断更新的状态变量。这些是动态规划变量, 每个变量定义了所有模式和马尔可夫链的成对组合。例如 $\alpha_t(c_i, g_t)$ 是 t 时刻对应于模式 c_i 和马尔可夫链 g_t 的前馈传播动态规划变量的值。在公式3-17和3-20中, 在每个时刻状态都进行更新, 这更新是通过将包含之前信息的马尔可夫跃迁矩阵和来自由下至上/由上至下的证据相结合得到的。

第四章 基于时间层级记忆模型的图像识别

在第四章中已介绍过节点在无噪声和有噪声情况下学习和推理的方法，由于无噪声属于理想情况，而在实际识别应用中基本处于有噪声情况，所以这里在对网络参数设计时假设处于有噪声的情况。

在有噪声情况下，如果不对输入进行预处理，那么有可能在level-1层产生大量的存储的模式，不仅占用大量存储资源，而且过大的计算量可能使学习无法实现。另外由于模式的类别过多，在学习后也无法体现出模式间的时间相关性，所以即使通过大量计算实现学习也不乏达到理性的识别效果。

预处理的算法基于输入的模式属于同一向量空间，空间中的不同模式使用欧几里德距离来判断模式间空间的相似性。欧氏距离彼此接近的模式会被划如同一个集群当中。节点使用 k 均值聚类算法使输入的模式聚集成 k 个集群， k 是节点的参数。 k 均值算法首先从输入中随机选取 k 个向量作为集群中心，当输入新的向量时就会被划入集群中心与它的欧氏最为接近的集群当中。然后根据新加入的向量重新计算集群中心，随着加入的向量越来越多，集群中心的变化量会减小直至稳定。

在感知/推理阶段，由于输入的模式不会与节点中存储的某个模式完全相同，因此需要求出输入模式与节点中已存储的每个模式的相似性。用 d_i 表示节点中存储的 i_m 模式与输入模式间的距离，距离越远相似性越低。输入模式与存储模式的相似度符合高斯分布，可使用 $e^{-d_i/\sigma}$ 计算，这里 σ 是节点的一个参数。对每个存储的模式进行这样的计算能够得到输入模式与马尔可夫图中各个顶点的相似性。

4.1 对 USPS 数据集的识别

USPS图集是通过收集得到的Buffalo地区信件上的手写数字，共包含9298个图，其中分为7291个训练图和2007个识别图，每幅图为16x16像素大小，图4-1显示了USPS图集的一些示例。

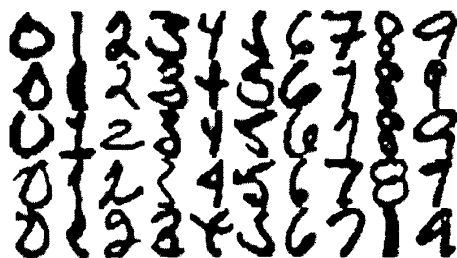


图 4-1 USPS 数据集示例

对于2007个识别图的人工识别准确率为97.5%^[30]。几种主要分类器关于USPS图像的识别率如表4-1所示。

表 4-1 几种主要方法的识别率

方法	识别率
相关向量机	94.9%
神经网络	95.8%
不变向量机	97%
KNN 分类器	94.4%
混合密度	92.8%
支持向量机	95.7%

因为USPS图像大小为16x16像素，所以我们构建的用于USPS识别的网络包含4层，处于最底层的level-1层包含有64个节点，8x8排列，每个节点对应USPS图的2x2个像素，之上的level-2层有16个节点，4x4排列，每节点对应2x2个level-1节点，以此类推，顶层有1个节点。

由于USPS为二值图像，所以直接将图像信息作为输入进入网络，节点之间的传播方式选择sumProp方式。我们在实验时首先采用1000幅图作为训练样本进行实验，观察识别率，如表4-2所示；之后采用7291幅全体训练样本进行实验，观察识别率，如表4-3所示。

表 4-2 1000 个训练样本实验

σ	k level-1	k level-2	k level-3	识别率
100	200	80	40	76.0%
50	200	80	40	81.3%
0.5	200	80	40	85.5%
0.5	300	100	48	88.1%
0.5	350	150	50	87.9%
0.25	300	100	48	88.6%
0.2	300	100	48	89.0%
0.15	300	100	48	89.6%
0.1	300	100	48	89.7%

表 4-3 全体训练样本实验

σ	k level-1	k level-2	k level-3	识别率
100	300	100	48	82.2%
50	300	100	48	84.6%
0.5	300	100	48	91.0%
0.25	300	100	48	93.3%
0.2	300	100	48	94.9%
0.15	300	100	48	95.6%
0.1	300	100	48	95.6%

从表4-2可以看出节点中的 k 并不是越大越好,在超过一定数量后不仅计算量增加,识别率反而会有所下降。 e^{-d^2/σ^2} 中的 σ 作为识别阶段判断模式相似性的参数,在开始阶段随着数值的减小识别率提高,在到达一定值以后识别率稳定,不再明显减小。在初始阶段使用1000图的样本以较快的速度进行较多的实验,能够得到参数的合理范围,再使用7291个全体训练样本进行实验,得到较为准确的识别率。通过与之前几类方法相比较,虽然此种方法得到的识别率并非最高,但结果还是令人满意的。

4.2 对 ALOI 数据集的识别

ALOI (Amsterdam Library of Object Images)是包含1000类物体的彩色图集,是专为科学研究之用搜集的。针对不变性识别研究,物体的视角、照明角度、照明颜色系统地变化。每类物体有超过100张图片,数据集图片总数超过110,250张。在我们的实验中使用其中的视角变化部分,并且简化为使用灰度图像。图像尺寸为192x144,每类物体包含72张图片。示例图像如4-2所示。

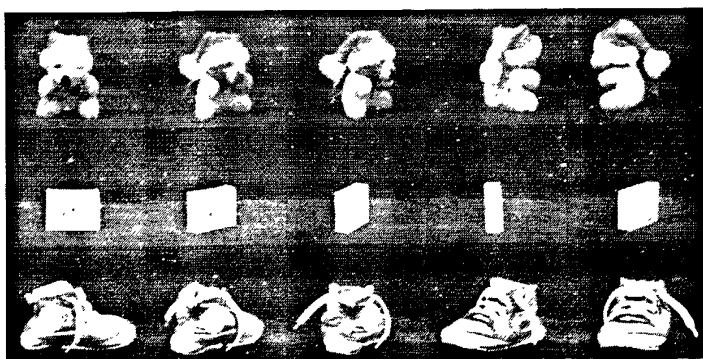


图4-2 ALOI中部分物体的视角变换

与USPS图集相比,ALOI要复杂的多,并且图片尺寸也更大。针对ALOI图集我们构建了7层的网络,并将图片尺寸变化为128x128像素,此外,对于输入网络的图像使用Garbor

滤波进行预处理，得到图像的轮廓线条，再输入进网络。

由于ALOI共有1000类物体，所以我们在实验时只随机选取部分类别，并对不同类别数的实验加以比较，每一类物体使用72幅图像作为识别样本。为简化考虑网络中的参数 k 和 σ 使用USPS实验得到的最优设置。实验结果如表4-3所示。

表4-3 ALOI识别实验

类别数 训练样本数	2	4	8	16	32
1	71.3%				
2	85.5%	70.9%	57.9%	31.6%	
4	93.6%	88.0%	77.4%	58.0%	53.5%
6	100.0%	94.3%	89.1%	73.3%	67.7%
9		100.0%	100%	86.9%	79.4%
12				100.0%	90.0%

从上表中可以看出识别率随着类别的增加明显下降，说明还存在着需要解决的问题。但是同时当类别较少时，增加训练样本可以得到很好的效果实现较高的识别率。

4.3 本章小结

作为一种新的模型，层级时间记忆模型在解决图像识别问题是的效果是值得肯定的。它的优势在于其原理符合生理学上对于大脑皮层的研究，同时由于大脑使用同样的算法解决不同种类的问题，使得这一模型也可以很容易的应用在其他领域中。但是它的缺点在于由于具有一定程度的通用性，所以对于视觉问题缺乏针对性，暂时还无法达到最高的识别率。

第五章 时间层级记忆模型中的改进聚类方法

基于哺乳动物大脑皮层生理特征建立的层级时间记忆模型，使用了层级结构模仿大脑皮层的逻辑层级，有助于对学习样本特征的归纳记忆，层级化的结构能够存储大量样本特征。同时在学习过程中使用时间信息作为参考，由于除最高层之外网络的其他层次节点的学习都属于无监督学习，因此将时间的临近性作为分类的重要特征，并且在物体的不变性识别过程中时间信息也非常重要。这是基于这样一个假设，即“在时间上临近发生的证据来自于同一物体”。这一点已经在视觉领域得到认可，人类通过视觉对于物体的认识和识别都是在对物体连续观察的基础上进行的。

无论是大脑皮层还是时间层级记忆模型，在其中每个层次中都存在着对于模式的聚类，而这一模型所采用的方法可能就是大脑皮层所使用的方法。在此模型中，首先对于输入模式进行聚类，将无限（或有限但因数量巨大无法计算）量的输入模式聚类为合理数量的模式并将聚类中心存储在节点中。存储模式稳定后，使用在训练中通过连续观察得到的模式间的时间相近性信息对已有的存储模式再次聚类，得到时间分组。高层节点将接收的子节点时间组信息作为输入模式，再进行以上步骤。由于基于“时间上临近发生的证据来自于同一物体”，因此子节点在向父节点传送时间分组信息时已经对子节点中的模式进行了很大程度上的归纳，并且由于参考了时间信息，可以忽略物体变形导致的模式特征的巨大变化，这对于不变性识别的作用是很重要的。

5.1 时间层级记忆模型中使用的聚类方法

此模型中主要在两个关键处使用了聚类方法，一是在 4.4.1 中提到的有噪声学习情况下使用欧几里得距离作为样本距离对输入模式进行聚类，这里使用了 k 均值的聚类算法，因在实际应用中基本都是噪声学习，所以此聚类基本上都会使用到。另一个是在 4.3.1 中提到的节点的学习中使用已得到的转换概率作为样本距离对存储模式进行聚类，最终得到时间分组，这里使用的是分层聚类算法。

5.1.1 k 均值方法

[1] 假设要聚成 k 个类。由人为决定 k 个类中心 $z_1(0), z_2(0), \dots, z_k(0)$;

[2] 在第 k 次叠代中，样本集 $\{z\}$ 用如下方法分类：

对所有 $i=1, 2, \dots, K, i \neq j$;

若 $\|z - z_j(k)\| < \|z - z_i(k)\|$ ，则 $z \in S_j(k)$;

[3] 令由[2]得到的 $S_j(k)$ 的新的类中心为 $Z_j(k+1)$;

$$\text{令 } J_j = \sum_{Z \in S_j(k)} \|Z - Z_j(k+1)\|^2 \text{ 最小, } j=1, 2, \dots, K;$$

$$\text{则 } Z_j(k+1) = \frac{1}{N_j} \sum_{Z \in S_j(k)} Z, N_j \text{ 为 } S_j(k) \text{ 中的样本数};$$

[4] 对于所有的 $j=1, 2, \dots, K$, 若 $Z_j(k+1)=Z_j(k)$, 则终止。否则 goto [2]。

开始时 K 的选择: 最初选择哪些样本作为中心, 将对叠代产生影响。多次叠代, 多次修正。

5.1.2 分层聚类算法

算法的思路是寻找“距离”最近的两个样本结合。

1. 有 N 个样本的集合 $Z_i = \{Z_1, Z_2, \dots, Z_N\}$;
2. 若想要聚成 K 个类:
 - [1] $k = N, C_i = \{Z_i\}, i=1, \dots, N$;
 - [2] If $k=K$ then END;
 - [3] 找到 C_i 与 C_j 之间的距离 $d(C_i, C_j)$ 最小的一对;
 - [4] C_i 和 C_j 合成一个类 C_i , 并计算新的 C_i 的中心;
 - [5] 去除 C_j , $k=k-1$, goto [2];

类间距离 $d(C_i, C_j)$;

1. 类中心间距: $d_1 = \|M_i - M_j\|$, 其中 $M_i = \frac{1}{n_i} \sum_{Z \in C_i} Z$, n_i 是属于 C_i 的样本数;
2. 靠得最近的样本: $d_2 = \min_{Z_i \in C_i, Z_j \in C_j} \|Z_i - Z_j\|$;
3. 离得最远的样本: $d_3 = \max_{Z_i \in C_i, Z_j \in C_j} \|Z_i - Z_j\|$;
4. 类间平均距离: $d_4 = \frac{1}{n_i n_j} \sum_{Z_i \in C_i} \sum_{Z_j \in C_j} \|Z_i - Z_j\|$;

距离计算次数 $C_N^2 = N(N-1)/2$ 。组合 $C_{N-1}^2, C_{N-2}^2, \dots$ 。

5.2 对输入模式的聚类算法进行改进

我们使用几种常见的聚类算法替换在有噪声学习情况下对于输入模式进行聚类的算法, 并通过实验验证算法的可行性。

5.2.1 最简单聚类方法

(1) 算法描述:

参考相似性尺度（距离）阈值，不需要事先给定 κ 。

有 N 个样本， $Z_i = \{z_1, z_2, \dots, z_N\}$ ；

给定一个阈值 τ ；

任取一个样本，例如 z_1 ，把 z_1 作为第一个类的中心， $z_1 = z_1$ ；

依次取 $z_i, i=1, 2, 3, \dots, N$ ，计算 z_i 与 z_1 的距离 D_{i1} ；

若 $D_{i1} \leq \tau$ ，则判定 z_i 属于 z_1 为中心的那个类；

若 $D_{i1} > \tau$ ，则把 z_i 作为新的类中心 z_2 。

对剩下的样本 z_i 分别计算与 z_1, z_2 的距离 D_{i1}, D_{i2} ；

若其中较小者 $\leq \tau$ ，则判定 z_i 属于较小的那一类；

否则，就把 z_i 作为新的一个类的中心 z_3 ；

如此继续，直至对全体样本做完处理。

算法特点：不需要事先决定类数。适用于类内距离小，类间距离大的情况。否则结果与取样本的顺序有关，亦与 τ 相关。

(2) 复杂度分析

对于第 n 个输入的模式，最多进行 $n-1$ 次比较，因此时间复杂度为 $O(n^2)$ 。

(3) 实验验证

实验仍然使用 USPS 数据集进行，与 κ 均值方法相比，识别率有所下降，但因计算量减少训练时间明显缩短。具体如表 5-1 所示。

表 5-1 κ 均值聚类与简单聚类对比

σ	κ 均值识别率	κ 训练时间	τ	简单聚类识别率	简单聚类训练时间
0.25	93.3%	3h4m41s	1	89.7%	1h22m47s
0.2	94.9%	3h39m17s	1	91.1%	1h58m20s
0.15	95.6%	3h50m40s	1	92.9%	2h26m6s
0.1	95.6%	4h11m4s	1	93.6%	3h5m17s

5.2.2 最大距离样本法

(1) 算法描述：

算法思路取尽可能离得远的样本做中心。

有 N 个样本， $Z_i = \{z_1, z_2, \dots, z_N\}$ 。

[1] 任取一个样本，例如 z_1 ，把 z_1 作为第一个类的中心， $z_1 = z_1$ ；

[2] 从集合 Z_i 中找出到 z_1 距离最大的样本作为 z_2 ；

[3] 对 Z_i 中剩余样本 z_i ，分别计算到 z_1, z_2 的距离。令其中较小的那个为 D_{iz} ；

[4] 计算 $\max_z \{D_{iz}\}$ 。若其值大于某一计算值或给定阈值，则取此 z_i 为新的类中心。计

算值可取：大于等于 z_i 和 z_j 间距离的 n/m 倍 ($1/2 \leq n/m < 1$)；

[5] 重复同样的处理，直到再也找不到符合条件的新的类中心；

[6] 把剩余样本分配到离它最近的那个中心所属的类。

缺点：与首先选取哪个样本有关。

(2) 复杂度分析

对于 $n-i$ 个样本，每个进行 $(i-1)+(n-i)=n-1$ 次比较运算，因此时间复杂度符合 $O(n)$ 。

(3) 实验验证

聚类原理与简单聚类基本相似，而 n 和 m 的取值具有很大随意性，因此暂取 $n=2$ ， $m=3$ ，试验结果如表 5-2 所示。

表 5-2 k 均值聚类与最大距离样本聚类对比

σ	k 均值识别率	k 均值训练时间	最大距离样本识别率	最大距离样本训练时间
0.25	93.3%	3h4m41s	89.9%	1h58m6s
0.2	94.9%	3h39m17s	91.7%	2h19m37s
0.15	95.6%	3h50m40s	92.8%	2h47m9s
0.1	95.6%	4h11m4s	93.2%	3h30m28s

可以看到最大距离样本相比于简单聚类识别率有所下降，这可能与初始中心的选取有关。另外计算量有所增加使得最大距离样本聚类法比简单聚类法训练时间延长。

5.3 对于按时间分组过程中聚类算法的改进

通过实验我们决定选择简单聚类法作为输入模式的聚类方法，接下来要分别使用 K 均值和叠代自组织聚类法替换分层聚类法，通过试验检验是否能够显现更好的效果。

5.3.1 K 均值方法

(1) 算法描述见 5.1.1

(2) 复杂度分析

由于已经预先分为 K 个类，因此第 k 次叠代时，一个模式需要进行 K 次比较，之后最多进行 n 次的求和计算，所以时间复杂度为 $O(kKn)$ 。

(3) 实验验证

K 均值的明显优势在于具有线性时间复杂度，但其劣势同样明显，因为事先人为确定类别数，并且类别中心随机确定，因此 K 均值方法一般只能达到局部最优，很难实现全局最优。

表 5-3 分层聚类法与 k 均值法的比较

σ	分层聚类识别率	分层聚类训练时间	k	k 均值识别率	k 均值训练时间
0.25	89.7%	1h22m47s	32	80.9%	59m45s
0.2	91.1%	1h58m20s	32	85.7%	1h20m9s
0.15	92.9%	2h26m6s	32	88.0%	2h5m43s
0.1	93.6%	3h5m17s	32	91.5%	2h19m31s

5.3.2 叠代自组织 (ISODATA) 聚类法

(1) 算法描述:

算法思路给定一些大致参数。

原则: ① 样本数太少的类 - 取消;

② 类内离散太大的类 - 分裂;

③ 距离近的类 - 合并。

1) 给一些参数:

K : 期望分类个数的大致范围;

θ_k : 一个类内的最少样本数;

θ_s : 关于类内分散程度的参数;

θ_c : 关于类间距离 (最小) 的参数;

L : 每次叠代允许合并的类数;

I : 允许叠代的最大次数。

2) 适当选取类中心 $\{z_1, z_2, \dots, z_{N_c}\}$, N_c : 类数。

2)' 分配样本。如果有 $\{i=1, 2, \dots, N_c\}$;

$\|z - z_j\| < \|z - z_i\|$, 则 $z \in S_j$, $j=1, 2, \dots, N_c$ 。

3) 如果 S_j 类样本数 $N_j < \theta_k$, 则取消 S_j 类。 $N_c = N_c - 1$, goto 2)'。

4) 重新计算各类中心 $z_j = \frac{1}{N_j} \sum_{z \in S_j} z$, $j=1, 2, \dots, N_c$ 。

5) 计算类 S_j 内平均距离 $\bar{D}_j = \frac{1}{N_j} \sum_{z \in S_j} \|z - z_j\|$, $j=1, 2, \dots, N_c$ 。

6) 对全体样本求类内距离平均值 $\bar{D} = \frac{1}{N} \sum_{j=1}^{N_c} N_j \cdot \bar{D}_j$, $N = \sum_{j=1}^{N_c} N_j$ 。

7) [a] 如果叠代次数 $\geq I$, 则转向 11) (合并);

[b] 若 $N_c \leq K/2$, 则转向 8) (分裂);

[c] 若偶数次叠代或 $N_c \geq K/2$, 则转向 11) (合并);

8) 计算各类中各分量的标准差。

$$\sigma_{ij} = \sqrt{\frac{1}{N_j} \sum_{z \in S_j} (x_{ik} - z_{ij})^2}, \quad i=1, 2, \dots, n, \quad j=1, 2, \dots, N_c, \quad k=1, 2, \dots, N_j;$$

x_{ik} 为 $Z \in S_j$ 的第 i 个分量, z_{ij} 为 Z_j 的第 i 个分量;

σ_{ij} 为第 j 类第 i 个分量标准差;

9) 找到各类的标准差最大的分量。

$$\sigma_{j, \max} = \max\{\sigma_{1j}, \sigma_{2j}, \dots, \sigma_{nj}\}, \quad j=1, 2, \dots, N_c;$$

10) 分裂: 条件1. $\sigma_{j, \max} > \theta_s$ 且 $\bar{D}_j > \bar{D}$ 且 $N_j > 2(\theta_k + 1)$ 。

条件2. $\sigma_{j, \max} > \theta_s$ 且 $N_c < K/2$;

若满足两条件之一, 则分裂 S_j ;

(a) 建立 Z_j^+ 和 Z_j^- , 2个新的类中心, $N_c = N_c + 1$;

其中 Z_j^+ 和 Z_j^- 是沿着 $\sigma_{j, \max}$ 轴, 在原来的 Z_j 位置上, 分别加上和减去一个数 $k\sigma_{j, \max}$ ($0 < k \leq 1$)。 k 是经验值。

(b) goto 2) (分配样本)。

11) 计算所有各类中心的相互距离 $D_{ij} = \|Z_i - Z_j\|$, $i=1, 2, \dots, N_{c-1}$, $j=i+1, 2, \dots, N_c$ 。

12) 对于比 θ_c 小的 D_{ij} 从小到大排队。假定为:

$$D_{i_1 j_1} \leq D_{i_2 j_2} \leq \dots \leq D_{i_l j_l};$$

13) 按 $i=1, 2, \dots, L$ 的顺序, 把 $D_{i_l j_l}$ 对应的 Z_{i_l} 和 Z_{j_l} 合并。

$$Z_i^* = \frac{1}{N_{i_l} + N_{j_l}} [N_{i_l} Z_{i_l} + N_{j_l} Z_{j_l}], \quad N_c = N_c - 1$$

计算 $D_{i_l j_l}$ 时的 Z_i , Z_j , 若至少其中一个是在本次叠代中合并取得类中心, 则越过此项。

14) 若叠代次数 ≥ 1 , 或参数无改变, 则终止。

否则 goto 2) ', 需要时可返回1) 修改参数。

(2) 复杂度分析:

叠代自组织在分配样本之后计算类中心、类内平均距离、样本总平均距离等, 包括发生分裂与合并情况下的计算, 总体符合 $O(n^2)$ 时间复杂度, 但计算量较大。

表 5-4 分层聚类法与 ISODATA 法的比较

σ	分层聚类识别率	分层聚类训练时间	ISODATA 识别率	ISODATA 训练时间
0.25	89.7%	1h22m47s	92.4%	2h45m57s
0.2	91.1%	1h58m20s	93.9%	3h11m7s
0.15	92.9%	2h26m6s	95.1%	3h48m8s
0.1	93.6%	3h5m17s	95.4%	4h25m13s

(3) 实验验证:

ISODATA 算法通过调整样本所属类别完成样本的聚类,而且自动地进行类别的“合并”和“分裂”,得到的分类数也比较合理,从而实现了较高的识别率。另一方面可能由于模式之间的距离相差不明显会导致大量的“合并”和“分裂”计算,使得训练时间大大增加。如表 5-4:

5.4 本章小结

本章分别使用了简单聚类法、最大距离样本法、 K 均值法和叠代自组织法替换原模型中使用的 K 均值法和分层聚类法。分别是在输入模式聚类阶段使用简单聚类法、最大距离样本法与原有的 K 均值法相比较;在按转换概率计算划分时间组阶段使用 K 均值法和叠代自组织与原有的分层聚类法相比较。6 中聚类方法各有其优点与劣势,通过实验进行了比较。可能由于了解不够深入,实验中的参数设置和初始中心优化等还存在问题,导致实验结果可能出现偏差。但通过这些实验可以简单比较不同的聚类法,为模型的改进提供一些参考。实验中可以发现在输入模式聚类阶段较适用简单聚类法,在基本保持识别率的条件下提高了训练速度;在按时间分组阶段原始的分层聚类法具有相对优势,能够在识别率和训练速度上保持平衡。如果为追求识别率可以牺牲训练时间的话也可以使用叠代自组织法。另外通过不同的实验可以推测在两阶段分别使用 K 均值法和叠代自组织法应该会得到最高的识别率,当然也相应的需要最多的训练时间。

第六章 结束语

6.1 本文主要工作

本论文研究并改进一种全新的解决图像识别问题的方法，从大脑皮层的生理学研究角度出发，研究基于理解皮层结构和运行机制而构建的贝叶斯网络并加以改进，尝试解决解决图像识别中的一般性问题。通过研究与实践表明这一模型具有实践意义，能够应用到模式识别的应用当中。之后我们把此模型现在的算法加以改进，在模型的关键部位尝试分层聚类、最大距离样本法、K 均值、叠代自组织等多种聚类方法并通过试验加以验证，在模型通用性的基础上加强其识别率和学习效率，得到了使用不同方法测试的数据，为今后模型的进一步改进提供了参考。

6.2 有待进一步研究的内容

此模型较新，另外由于研究力量的限制，研究进行的没有更加深入，有一些内容还有待进一步研究，主要包括：

1. 研究更多的聚类算法对此模型的改进，以期能够更加显著提高识别的准确率和学习与识别的速度。
2. 通过研究更多的大脑生理结构的科研成果，将可能影响视觉识别的皮层机构和机制加入模型中，以此验证是否能提高模型的识别性能。

参考文献

- [1] 王秀珍. 图像识别技术浅论. 内蒙古电大学刊, 2008年第8期
- [2] 王耀南, 李树涛, 毛建旭. 计算机图像处理与识别技术. 北京:高等教育出版社, 2001. 6 ISBN 7-04-009468-1
- [3] 林春漪. 基于混合贝叶斯网络的医学图像语义建模及其检索的研究. 华南理工大学博士学位论文, 2006 6
- [4] 余东峰, 孙兆林. 基于贝叶斯网络不确定推理的研究. 微型电脑应用, 2004, 20(8):
- [5] 张磊. 基于贝叶斯网络的步态识别. 北京科技大学硕士学位论文, 2006
- [6] 孙即祥. 模式识别中的特征提取与计算机视觉不变量. 北京:国防工业出版社, 2004
- [7] 吴志勇, 蔡莲红. 基于动态贝叶斯网络的音视频双模态说话人识别. 计算机研究与发展, 43(3): 470-475, 2006
- [8] 赵玉霞, 王克如等. 贝叶斯方法在玉米叶部病害图像识别中的应用. 计算机工程与应用, 43(5):193, 2007
- [9] 陈菲. 基于贝叶斯理论的超声肝图像识别系统研究. 微计算机信息, 2007年第23卷第7-3期
- [10] 汪国有, 邹玉兰. 基于贝叶斯数据融合的多尺度目标识别. 华中科技大学学报, 2003年11月, 第31卷第11期
- [11] James Rehg, Kevin Murphy, Paul Fieguth. Vision-Based Speaker Detection Using Bayesian Networks. In Proceedings of Conference on Computer Vision and Pattern Recognition, 1999, 2:110-11
- [12] Henry Schneiderman. Learning a Restricted Bayesian Network for Object Detection. Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004, 639-64
- [13] Hawkins, J. and Blakeslee, S. 2004. *On Intelligence*. New York, NY: Times Books.
- [14] J. Pearl, *Probabilistic Reasoning in Intelligent Systems* (Morgan Kaufmann, San Mateo, Calif., 1988).
- [15] J. Yedidia, W. T. Freeman, and Y. Weiss, Understanding belief propagation and its generalization, presented at the International Joint conference on Artificial Intelligence (IJCAI 2001), Seattle, Washington, August 4-10, 2001.
- [16] Shaul Hochstein and Merav Ahissar View from the Top: Hierarchies and Reverse Hierarchies in the Visual System Neuron, Vol. 36, 791-804, December 5, 2002
- [17] Maximilian Riesenhuber and Tomaso Poggio. Hierarchical models of object recognition in cortex. Nature Neuroscience, 2(11):1019-1025, November 1999.
- [18] T. S. Lee, C. Yang, R. Romero, and D. Mumford, "Neural activity in early visual cortex reflects behavioral experience and higher order perceptual saliency," Nat. Neurosci. 5, 589-597 (2002).
- [19] Tai Sing Lee and David Mumford. Hierarchical Bayesian inference in the visual cortex. J Opt Soc Am A Opt Image Sci Vis, 20(7):1434-1448, Jul 2003.
- [20] Suzanna Becker. Implicit learning in 3D object recognition: The importance of temporal context.

- Neural Computation, 11(2):347-374, February 1999.
- [21] Peter Foldiak. Learning invariance from transformation sequences. *Neural Computation*, 3(2):194-200, 1991.
- [22] Simon M. Stringer and Edmund T. Rolls. Invariant object recognition in the visual system with novel views of 3D objects. *Neural Computation*, 14(11):2585-2596, November 2002.
- [23] George D., and Hawkins, J. 2005. A Hierarchical Bayesian Model of Invariant Pattern Recognition in the Visual Cortex. In *Proceedings of the International Joint Conference on Neural Networks 2005*. Montreal, Canada: International Neural Network Society
- [24] Jeff Hawkins. Why Can't A Computer Be More Like A Brain? Or What To Do With All Those Transistors? *IEEE International Solid-State Circuits Conference*, 2008
- [25] Jeff Hawkins and Dileep George. *Hierarchical Temporal Memory: Concepts, Theory, and Terminology*. Whitepaper, Numenta Inc.
- [26] Kenneth L. Rice, Tarek M. Taha, Christopher N. Vutsinas. Scaling analysis of a neocortex inspired cognitive model on the Cray XD1. Springer Science Business Media, LLC 2008
- [27] Dileep George and Bobby Jaros. *The HTM Learning Algorithms*. Whitepaper, Numenta Inc.
- [28] Saulius J. Garalevicius. Memory-Prediction Framework for Pattern Recognition: Performance and Suitability of the Bayesian Model of Visual Cortex. *American Association for Artificial Intelligence*, 2007
- [29] Perlis, D. 2005. Hawkins on Intelligence: Fascination and Frustration. *Artificial Intelligence* 169(2005):184-191.
- [30] Feldman, J. A. 2005. On Intelligence as Memory. *Artificial Intelligence* 169(2005):181-183.
- [31] R. O. Duda and P. E. Hart. *Pattern Classification and Scene Analysis*. John Wiley and Sons, New York, 1973.
- [32] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.
- [33] V B Mountcastle. An organizing principle for cerebral function: the unit module and the distributed system. In *The Mindful Brain*. MIT Press, 1978.
- [34] L von Melchner, S L Pallas, and M Sur. Visual behaviour mediated by retinal projections directed to the auditory pathway. *Nature*, 404(6780):871-876, 2000.
- [35] Tom M. Mitchell. *The need for biases in learning generalizations*. Technical Report CBM-TR-117, New Brunswick, New Jersey, 1980.
- [36] William H. Calvin and George A. Ojemann, *Inside the Brain: Mapping the Cortex, Exploring the Neuron*, New American Library. 1980.
- [37] Brian A Wandell, Serge O Dumoulin, and Alyssa A Brewer. Visual field maps in human cortex. *Neuron*, 56(2):366-383, 2007.
- [38] Rodney J. Douglas and Kevan A. Martin. Neuronal circuits of the neocortex. *Annual Review of Neuroscience*, 27:419-451, 2004. LR: 20041117; JID: 7804039; RF: 176; ppublish.
- [39] Vernon B. Mountcastle. The columnar organization of the neocortex. *Brain*, 120(4):701-22, 1997.
- [40] D. J. Felleman and D. C. Van Essen. Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex*, 1(1):1-47, 1991.
- [41] Herbert A. Simon. *The Sciences of the Artificial*. MIT Press, Cambridge, Massachusetts, 1981.
- [42] Jay Hegde and Daniel J Felleman. Reappraising the functional implications of the primate visual anatomical hierarchy. *Neuroscientist*, 13(5):416-421, 2007.
- [43] David H. Wolpert and William G. Macready. No free lunch theorems for optimization. *IEEE Trans. on Evolutionary Computation*, 1(1):67-82, 1997.
- [44] Peter A. Bannister. Inter- and intra-laminar connections of pyramidal cells in the neocortex. *Neurosci Res*, 53(2):95-103, 2005.

- [45] Jeff Hawkins and Sandra Blakeslee. *On Intelligence*. Henry Holt and Company, New York, 2004.
- [46] Dileep George and Bobby Jaros. *The HTM Learning Algorithms*. www.numenta.com.
- [47] George, D., J. Hawkins.: A hierarchical Bayesian model of invariant pattern recognition in the visual cortex. In: *Proceedings of the International Joint Conference on Neural Networks (IJCNN-05)*, 2005
- [48] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Francisco, California, 1988.
- [49] Stephan C. Johnson. Hierarchical clustering schemes. *Psychometrika*, 32:241-254, 1967.
- [50] Bromley and Sackinger. Neural-network and K-nearestneighbor classifiers. Technical Report 11359-910819-16TM, AT&T, 1991.

发表论文和科研情况说明

发表的论文：

- [1] Lei Wang, Xian-Bin Wen, Xu Jiao, Jian-Guang Zhang, "Object Recognition using a Bayesian Network imitating Human Neocortex", CISP2008, 2009年10月

参与的科研项目：

- 1.国家自然科学基金：“多尺度概率图模型与 SAR 图像分类的研究”（60872064）
- 2.天津自然科学基金：“多尺度图模型与遥感图像分类的研究”（08JCYBJC12300）

致 谢

本论文的工作是在我的导师温显斌教授的悉心指导下完成的，温显斌教授严谨的治学态度和科学的工作方法给了我极大的帮助和影响。在此衷心感谢三年来温显斌老师对我的关心和指导。

温显斌教授悉心指导我们完成了实验室的科研工作，在学习上和生活上都给予了我很大的关心和帮助，在此向温显斌老师表示衷心的感谢。

在实验室工作及撰写论文期间，张建光、焦旭等同学对我论文中的研究工作给予了热情帮助，在此向他们表达我的感激之情。

另外也感谢我的父母和女友，他们的理解和支持使我能够在学校专心完成我的学业。