

稀疏认知学习、计算与识别的研究进展

焦李成¹⁾ 赵进¹⁾ 杨淑媛¹⁾ 刘芳^{1),2)} 谢雯¹⁾

¹⁾(西安电子科技大学智能感知与图像理解教育部重点实验室、国际智能感知与计算联合研究中心 西安 710071)

²⁾(西安电子科技大学计算机学院 西安 710071)

摘 要 稀疏认知学习、计算与识别是近年来受到国际学术界广泛关注的学术前沿领域,这一新的学习、计算与识别范式将对机器学习、模式识别、计算智能以及大数据等领域的研究产生变革性的影响.为能更好地把握其发展规律,本文以生物视觉稀疏认知机理的研究进展为依据,通过对生物视觉稀疏认知机理的学习及建模,详细地评述了稀疏编码模型、结构化稀疏模型和层次化稀疏模型的研究进展,并阐释这 3 种模型之间的区别与联系;以及概述稀疏认知计算模型目标识别应用的研究状况,并举例给出了稀疏认知计算模型的具体成果;最后总结了这一新范式存在的一些问题及可进一步研究的方向,以期引起更多的研究同行对这一具有交叉性和发展性方向的关注.

关键词 稀疏编码;结构化稀疏;层次化稀疏;视觉皮层;稀疏认知学习;计算与识别;认知计算;机器学习;人工智能
中图法分类号 TP18 **DOI号** 10.11897/SP.J.1016.2016.00835

Research Advances on Sparse Cognitive Learning, Computing and Recognition

JIAO Li-Cheng¹⁾ ZHAO Jin¹⁾ YANG Shu-Yuan¹⁾ LIU Fang^{1),2)} XIE Wen¹⁾

¹⁾(Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education,
International Research Center for Intelligent Perception and Computation, Xidian University, Xi'an 710071)

²⁾(School of Computer Science and Technology, Xidian University, Xi'an 710071)

Abstract Sparse Cognitive Learning, Computing and Recognition is a widespread concern front filed by international academic circles in recent years, this new paradigm will have be a revolutionary impact on the fields of machine learning, pattern recognition, computational intelligence, big data etc. In order to better grasp the law of development, three forms of this new paradigm, such as sparse coding, structured sparse and hierarchy sparse, are reviewed in detail from the perspective of biological visual cognitive mechanism and model structure in this paper. In the meantime, outlined the application research of this computational model and exemplified some concrete achievements. Furthermore, some existing problems and research directions are discussed by combining with the current application research. In order to arouse interested researcher pay more attention to this new paradigm.

Keywords sparse coding; structured sparse; hierarchy sparse; visual cortex; sparse cognitive learning; computing and recognition; cognitive computing; machine learning; artificial intelligence

收稿日期:2014-12-23;在线出版日期:2015-05-17.本课题得到国家“九七三”重点基础研究发展规划项目基金(2013CB329402)、高等学校学科创新引智计划(B07048)、国家自然科学基金重大研究计划(91438201,91438103)及教育部“长江学者和创新团队发展计划”(IRT1170)资助.焦李成,男,1959年生,博士,教授,中国计算机学会(CCF)高级会员,主要研究领域为智能感知、图像理解等. E-mail: lchjiao@mail.xidian.edu.cn.赵进,男,1986年生,博士研究生,主要研究方向为稀疏表示与建模.杨淑媛,女,1978年生,博士,教授,主要研究领域为智能信号与图像处理、机器学习等.刘芳,女,1963年生,教授,主要研究领域为智能信息处理、模式识别等.谢雯,女,1989年生,博士研究生,主要研究方向为极化 SAR 图像分类.

1 引言

智能信息处理是当前信息科学理论和应用研究中的一个热点领域. 由于计算机网络及传感技术的发展, 社交网络、遥感及医学等诸多领域都面临着急剧增长的大数据. 面对数据的不断增长, 对数据处理工具的要求也越来越高, 人们希望在合理的时间内有效地从这种具有价值密度低且复杂结构的数据中获取其潜在的知识. 特别是近 20 年间, 科学家们为设计出有效的数据处理算法, 将目光聚焦到了智能的大脑系统建模上, 特别是作为该研究方向之一的生物视觉认知机理的建模受到了人工智能和模式识别领域的广泛重视.

众所周知, 信号表示是信号与信息处理领域中的一个核心问题. 稀疏表示是继小波分析, 多尺度几何分析之后, 被提出的又一类新的信号表示方法, 它通过基或字典中很少量的原子的线性组合来表示信号. 然而有关稀疏表示最初的思想却来源于生物视觉认知领域, 具体体现在稀疏的概念上, 它是 1959 年 Hubel 和 Wiesel^[1] 在研究猫的视觉条纹皮层上的细胞感受野时首次提出的, 结论为“初级视觉皮层(即 V1 区)上的细胞的感受野能够对视觉感知信息产生一种稀疏的响应, 即大部分神经元处于静息状态, 只有少数的神经元处于刺激状态”. 另外, 关于稀疏表示, 可以追溯到 1993 年法国数学家 Mallat 等人^[2] 基于小波分析提出的信号可以用一个过完备的字典来进行表示, 继而开启了稀疏表示的先河. 由于稀疏表示的思想新颖、方法独特, 使得它成为了一种重要的智能信息处理技术. 在实际信号处理任务中, 已发现许多自然信号在适当的变换下呈现稀疏性(即大多数变换系数为 0 或接近于 0, 仅有少数的变换系数不为 0). 因此, 为了有效利用稀疏性来实现数据的加工处理, 近 10 余年来, 基于生物视觉稀疏认知机理的数学建模成为了一个热门的研究方向, 主要是因为这种处理方式不但可以节省存储空间和降低时间复杂度, 而且还可快速地提取信号中的本质信息.

另外, 研究生物视觉皮层如何实现对外界刺激的稀疏响应与目标识别一直是视觉神经科学领域中的一个关键问题. 在过去的几十年里, 已有些科学家们利用视觉神经生理研究中所获得的实验数据, 并建立了一些计算模型, 在对这一关键问题的探索研究中, 取得了较好的成果. 如 1969 年 Willshaw 和 Buneman 等人^[3] 提出的基于 Hebbian 局部学习规

则的稀疏表示模型, 其中的稀疏表示可以使得记忆能力最大化, 进而有利于网络结构中联想机制的建立; 又如 1972 年 Barlow^[4] 给出了“稀疏性和自然环境的统计特性之间存在着某种相关性联系”的推论; 利用该推论, 1996 年 Olshausen 和 Field^[5] 提出了稀疏编码, 验证了自然图像经过稀疏编码后, 学习得到的基函数可以近似描述 V1 区上简单细胞的感受野的响应特性. 进一步, 随着生物视觉皮层中关于稀疏性研究的不断深入, 如从 V1 区上简单细胞的感受野特性逐渐发展到 V1 区上复杂细胞的感受野特性, 再到近年来依据自然场景的不同特征(如空间位置、形状、颜色、运动等)将视觉皮层分成不同的视觉通路实现并行处理的研究, 都已经取得了较好的研究成果; 对应着生物视觉稀疏认知机理方面所取得的研究进展, 稀疏认知计算模型也从稀疏编码^[5] 发展到了结构化稀疏模型^[6-8] 和判别性稀疏模型^[9] 等, 再到近两年来关注的层次化稀疏模型^[10-12]. 由于迄今为止, 生物视觉的稀疏认知机理还没有完全清楚, 仍需要不断地完善和发展, 再加上涉及计算机科学、神经认知科学和应用数学等诸多领域学科的交叉结合, 使得基于生物视觉稀疏认知机理的建模任务仍任重道远.

本文论述的稀疏认知学习、计算与识别范式可以理解为, 借鉴生物视觉的稀疏认知机理, 学习并完成该机理的数学建模, 进而通过得到的稀疏认知计算模型实现目标(如自然图像等)的识别. 这一范式主要包含了生物视觉的稀疏认知机理, 基于生物视觉稀疏认知机理的学习及数学建模, 和基于稀疏认知计算模型的目标识别这 3 个方面. 围绕着这 3 个方面, 本文将对稀疏认知学习, 计算与识别范式的研究进展进行综述.

在本文中, 第 2 节介绍生物视觉稀疏认知机理的研究进展; 第 3 节将给出基于生物视觉稀疏认知机理的学习和数学建模; 第 4 节详细地评述稀疏认知计算模型的结构, 算法研究和发展现状; 第 5 节将举例说明稀疏认知计算模型的目标识别的研究进展; 第 6 节将给出存在的问题及进一步研究的方向; 最后是本文的总结.

2 生物视觉稀疏认知机理的研究进展

2.1 生物视觉稀疏认知机理的生理实验依据

众所周知, 生物视觉皮层中的 V1 区在视觉信息处理中具有极其重要的作用, 但是从 20 世纪 60 年

代 Hubel 和 Wiesel 开创性的研究工作到 20 世纪 90 年代 Olshausen 和 Field 提出的 V1 区上简单细胞的稀疏编码理论, 这 30 余年从整体上对生物视觉皮层信息处理机理的了解还是比较少的. 为了探讨 V1 区对外界刺激是否采用了神经稀疏编码策略, 同时又为了避免计算模型设定的过多假设, 2008 年国内学者赵松年和姚力等人对该区进行了初步的功能性核磁共振实验, 通过给定两类视觉刺激图像, 每一类都具有大尺度特征和不同的细节, 实验结论为: 具有相同的轮廓与形状但细节不同的视觉刺激, 所引起 V1 区的活性模式是相似的, 即具有近似不变性. 本质上, 这是生物视觉皮层整合整体特征的体现, 即 V1 区对图像整体特征的同步化响应以及高级皮层区形状感知对 V1 区反馈的协同作用; 同时也是神经稀疏编码的体现^[13]. 另外, 这一结论与神经生物学家 Houweling 和 Brecht 等人^[14]在 2008 年《Nature》杂志上发表的文章类似, 都从生物视觉神经生理实验的角度有效支撑了神经稀疏编码的假说. 进一步的研究已证实, 神经稀疏编码原则贯穿于生物视觉皮层处理信息的多个阶段, 不只是在 V1 区.

同时, 神经生理学家通过对生物视觉皮层进行解剖、电生理、功能性核磁共振等手段, 已经将大脑皮层按功能分成了很多个区, 其中与视觉皮层有关的区有 20 多个. 目前, 从神经生理的角度, 研究较清楚的有初级视觉皮层(V1 区), 次级视觉皮层(V2 区), 高级视觉皮层(V4 区)和前/后下颞叶皮层(PIT/AIT 区)等; 而另一些还不是很清楚^[15], 如颞中回(MT)和前额皮层(PFC). 另外, 神经生理实验已表明, 生物视觉系统会根据自然场景中的不同特征将视觉皮层分成不同的通路并进行并行处理, 其中每一条通路为串行的等级结构. 最重要的两条通路是背侧视觉通路和腹侧视觉通路, 前者完成“在哪儿”的功能, 后者完成“是什么”的功能, 其中模式识别领域的学者最为关心的是腹侧视觉通路, 该通路采用 V1↔V2↔V4↔PIT/AIT↔PFC 的串行等级结构来进行信息处理(见图 1).

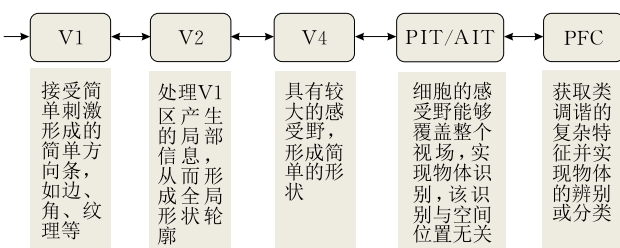


图 1 腹侧视觉通路的层次结构及其处理特点

通过分析外界场景中目标的形状、颜色和尺寸等静态特征后实现目标识别, 注意该层次结构中的层与层之间存在着前馈连接、反馈调节和水平侧向抑制的交互作用. 另外, 对生物视觉皮层感知机理的研究已表明, 视皮层不同区上的神经细胞的感受野具有不同的特性, 例如 V1 区上的简单细胞的感受野具有带通、方向和局部化特性; 而该区上的复杂细胞除了带通和方向特性外, 其感受野还具有一定的平移不变性. 下面, 依据文献^[16]简要地给出该通路上各视皮层区细胞的 3 种特性见表 1.

表 1 腹侧视觉通路各皮层神经细胞的特性

视皮层分区	感受野大小	刺激特征	尺度不变性
V1 区	1°~1.5°	边缘/线条	小
V2 区	3.2°	边缘/线条/角	小
V4 区	8°	中等复杂特征	较大
PIT 区	20°	中等复杂特征	较大
AIT 区	50°	视调谐复杂特征	约 2 倍缩放
PFC 区	很大	类调谐复杂特征	约 2 倍缩放

从表 1 中可以看出, 不同视皮层区上的神经细胞对特定形状的视觉图案有最佳的响应或偏好刺激, 这个可用感受野的术语来描述, 层次越高则感受野越大, 即信息处理是从局部到更大的区域. 有兴趣了解感受野的研究与进展, 请参考文献^[17]. 有关腹侧视觉通路各皮层细胞的其它特性(如方向和位置的不变特性等)及建立在该通路上的视觉计算模型的研究进展, 请参考文献^[16].

2.2 生物视觉稀疏认知机理的研究目的

将从神经稀疏编码定义^[18]为模拟生物视觉系统细胞感受野的一种人工神经网络^[19-21]方法, 可以看出生物视觉皮层的生理研究是非常重要和基础的. 因为利用生物视觉生理研究所获得的实验数据及合理的视觉皮层稀疏性假设, 可以建立相应的计算模型并借助计算机来验证科学家们对生物视觉稀疏认知机理的理解是否正确, 这也正是生物视觉稀疏认知机理研究的主要目的. 但这些模型^[16](例如 Visnet 模型、Neocognitron 模型、层次目标识别模型等)只是对视觉皮层信息处理机理的一种简化, 与真正的生物视觉系统的功能仍有很大的差距, 所以研究人员们期望建立一种具备预测性能的计算模型, 使得对生物视觉稀疏认知机理的研究可进行反馈与指导, 以便寻求建立更为合理的计算模型. 总之, 生物视觉稀疏认知机理的研究与视觉计算模型相结合是理解生物视觉稀疏认知机理的一种有效方法.

2.3 生物视觉稀疏认知机理的研究现状

随着神经生理技术的不断提高, 对生物视觉皮

层的研究也有了突飞猛进的进展. 代表性的工作如下, 1959 年 Hubel 和 Wiesel 对猫的视觉皮层进行研究并得到“V1 区上细胞的感受野能够对视觉感知信息产生一种稀疏表示”, 因这方面的出色工作他们获得了 1981 年的诺贝尔奖. 与此同时, 利用 V1 区的研究成果, 对生物视觉皮层按功能特性分区的研究也格外引人注目^[15], 1987 年 Field 等人提出 V1 区上的简单细胞可以产生自然图像的稀疏表示; 基于这个结论, 1988 年 Mitchison^[22] 明确提出了神经稀疏编码的概念; 之后牛津大学的 Rolls 等人^[23] 于 1990 年正式引用了神经稀疏编码这一概念, 通过对灵长类动物视觉皮层做电生理实验, 结果也进一步证实了视觉皮层对外界场景的刺激响应采用的是稀疏编码策略; 当 Hyvarinen 和 Hoyer^[24] 于 2001 年将这种刻画 V1 区上简单细胞的感受野特性延拓到该区的复杂细胞上时, 构造了由简单和复杂单元所共同形成的两层稀疏编码模型, 经验证该模型能较好的反应 V1 区上的复杂细胞的感受野特性.

生物视觉皮层如何识别复杂场景下的目标一直是生物视觉神经科学领域中的一个热点问题. 针对该问题, 科学家们结合生物视皮层生理研究的成果, 建立了视觉计算模型来对这种识别机理进行模拟并给出合理的解释. 最具代表性的工作有 1999 年美国麻省理工学院人工智能实验室的 Riesenhuber 和 Poggio^[12] 提出了层次目标识别模型(即 HMAX), 经验证该模型能够较为合理的解释哺乳动物腹侧视觉通路的信息处理机理; 随后 2007 年, Serre 等人^[10] 扩展了 HMAX 模型, 通过引入特征编码字典, 使得改进后的模型可以定量的模拟灵长类动物视觉皮层腹侧通路的信息处理机理; 进一步, 2014 年清华大学 Hu 等人^[11] 在 HMAX 模型的基础上, 通过引入稀疏正则的约束, 使得复杂场景下的目标识别任务在保证高性能的前提下, 其处理的速度, 较之前的 HMAX 模型提升了许多. 另外, 有兴趣了解更多关于生物视觉皮层中神经稀疏编码的研究历史, 请参考文献^[18].

3 基于生物视觉稀疏认知机理的学习与建模

基于生物视觉的稀疏认知机理, 并依据下面的研究脉络, 即从 V1 区简单细胞的感受野特性到该区复杂细胞的感受野特性, 再到腹侧视觉通路的机理; 来学习生物视觉稀疏特性并完成对应的稀疏认

知机理的建模, 即 V1 区上简单细胞、V1 区上复杂细胞和腹侧视觉通路的建模.

3.1 V1 区简单细胞的稀疏性学习与建模

根据神经稀疏编码假说, V1 区上的每个神经元对外界场景的刺激均采用了稀疏编码的形式来进行描述. 进一步, 为了理解 V1 区简单细胞的感受野特性, 1996 年 Olshausen 和 Field 沿着 Barlow^[4] 给出的“神经元的稀疏性和自然环境的统计特性之间存在着某种联系”的思路, 通过假设自然图像与基函数之间存在着线性关系, 建立如下的数学模型:

$$I(x, y) = \sum_i a_i \phi_i(x, y),$$

并首次采用编码系数的稀疏正则化约束, 得到如下的优化问题:

$$\min E = -[\text{Preserve Information}] - \lambda[\text{Sparseness of } a_i],$$

其中, 信息保真项和稀疏约束项分别为

$$[\text{Preserve Information}] = - \sum_{x, y} [I(x, y) - \sum_i a_i \phi_i(x, y)]^2,$$

$$[\text{Sparseness of } a_i] = - \sum_i S\left(\frac{a_i}{\sigma}\right),$$

其中: $I(x, y)$ 表示自然图像; $\phi_i(x, y)$ 为基函数; a_i 为编码系数; σ 为一个尺度常数; 函数 $S(x)$ 可以选择 $-e^{-x^2}$, $\log(1+x^2)$ 和 $|x|$ 等形式, 这些选择都可以促使得到的编码系数具有稀疏性, 即使编码系数具有较少的非零系数项.

基于此优化问题, Olshausen 和 Field 等利用自然图像 $I(x, y)$ 作为输入, 学习了基函数和其编码系数, 实验结果表明, 得到的基函数能够近似地反映 V1 区上简单细胞的感受野特性, 即带通、方向和局部化特性; 其中带通性指的是多分辨特性, 即能够对自然图像从粗分辨率到细分辨率进行连续地逼近; 方向性指的是基函数须具有各向异性; 局部化特性指的是基函数的时频局部化分析能力. 随后 1997 年他们又考虑了过完备基, 又称为字典, 并提出了过完备基的稀疏编码算法^[25].

3.2 V1 区复杂细胞的稀疏性学习与建模

神经生理研究发现, V1 区上复杂细胞的感受野与该区简单细胞的感受野特性大体上一致, 都具有严格的方向和带通特性, 不同的是复杂细胞的感受野具有局部的平移不变性. 为了研究 V1 区上复杂细胞的感受野特性, 2001 年芬兰学者 Hyvarinen 和 Hoyer 设计了一个两层的网络结构模型, 该模型是由简单细胞层和复杂细胞层组成(见图 2), 不仅要

求简单细胞层对外界刺激的响应是稀疏的,而且还要求复杂细胞层上的响应需具有空间局部稀疏特性,即在任意给定的时间内,简单细胞的非零响应应具有聚类特性或空间拓扑特性.模型的具体结构描述如图 2 所示.

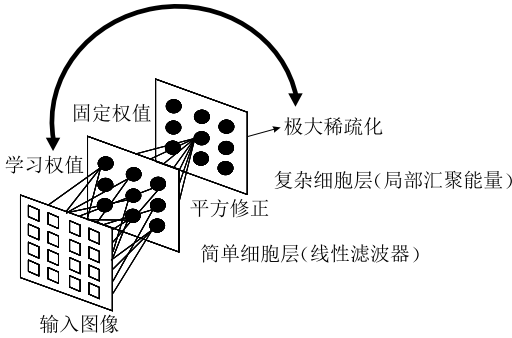


图 2 两层网络结构模型^[15]

给出 T 幅训练图像 $I_t(x, y), t=1, 2, \dots, T$, 并定义复杂细胞层上的输出响应为

$$c_{i,t} = \sum_{j=1}^n h(i, j) \langle \omega_j, I_t \rangle^2,$$

其中 $h(i, j)$ 为简单细胞层上的第 j 个细胞与复杂细胞层上的第 i 个细胞的连接或汇聚的权值, ω_j 为输入层到简单细胞层上的连接权值, 以及

$$\langle \omega_j, I_t \rangle = \sum_{x,y} \omega_j(x, y) I_t(x, y).$$

然后对输出响应进行极大化似然优化, 得到如下的优化问题:

$$\max \log L(I_1, I_2, \dots, I_T; \omega_1, \omega_2, \dots, \omega_T) = \sum_{t=1}^T \sum_{i=1}^n G(c_{i,t}),$$

这里的 G 为凸函数, 其中的 $h(i, j)$ 是事先固定的, 不需要通过训练数据来进行学习.

求解此优化问题, Hyvarinen 和 Hoyer 得到输入层到简单细胞层上的连接权值, 实验结果表明, 极大化似然优化问题等价于极大化稀疏约束的优化问题, 这点与 Olshausen 和 Field 等人的结论一致; 另外, 将求解的连接权值(可理解为滤波器)转化为对应的合成字典, 经过分析, 该字典具有空间拓扑特性或结构特性, 即反映了复杂细胞层上的响应具有的空间局部稀疏特性, 其中复杂细胞的稀疏响应特性通过下一高阶神经层(即 Contour Cell)来稀疏表示^[26]. 有关连接权值转化字典的方法参考文献^[24].

3.3 腹侧视觉通路的稀疏性学习与建模

基于生物视觉皮层腹侧通路的机理, 1999 年 Riesenhuber 和 Poggio 等人提出了 HMAX 模型, 它由对应于视觉皮层中简单细胞的 S 单元与复杂

细胞的 C 单元交替组成的前馈等级结构, 例如 4 层等级结构 $S1 \rightarrow C1 \rightarrow S2 \rightarrow C2$, 其中的 S 单元主要用于增加目标的选择特性, C 单元通过局部最大汇聚操作来增加对目标变换不变性的刻画. 该模型的优点是, 与腹侧视觉通路的机理大体上是一致的. 缺点是丢失了空间位置信息, 所以对目标不能实现精准定位, 并且完成信息处理的计算量偏大.

为了克服 HMAX 模型的缺点, 改进的方法很多, 下面本节具体地给出法国学者 Thieriault 等人^[27]于 2011 年提出的改进版 HMAX 模型, 其模型仍采用 4 层结构, 学习流程见图 3.

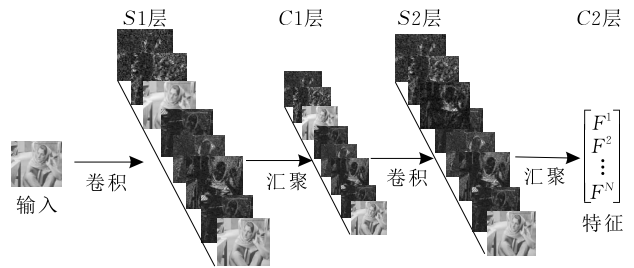


图 3 具有 4 层的 HMAX 模型的学习流程

根据图 3, 具体描述每一层上神经细胞的特性和操作如下:

S1 层: 主要模拟视觉皮层 V1 区及 V2 区上的简单细胞的感受野特性, 即带通、方向和局部化特性. 具体的操作是

$$S1: R^{m \times n} \rightarrow R^{m \times n \times \Theta \times K}$$

$$I(x, y) \rightarrow [S_{1,1}, S_{1,2}, \dots, S_{\Theta,K}]^T,$$

其中的 Θ, K 分别为多方向, 多尺度滤波器 $g_{\theta, \sigma}(x, y)$ 的方向个数和尺度个数, 即方向 $\theta_i \in (\theta_1, \theta_2, \dots, \theta_\Theta)$ 和尺度 $\sigma_j \in (\sigma_1, \sigma_2, \dots, \sigma_K)$; 另外

$$S_{1,i,j} = g_{\theta_i, \sigma_j}(x, y) \times I(x, y),$$

其中 $i \in [1, 2, \dots, \Theta]$ 和 $j \in [1, 2, \dots, K]$. 通常这里的滤波器选择为 Gabor 滤波器.

C1 层: 主要模拟视觉皮层 V1 区, V2 区上的复杂细胞的感受野特性, 即方向性, 带通性和局部的平移不变性. 具体的操作是

$$C1: R^{m \times n \times \Theta \times K} \rightarrow R^{o \times p \times \Theta \times K}$$

$$\begin{bmatrix} S_{1,1,1} \\ \vdots \\ S_{1,\Theta,K} \end{bmatrix} \rightarrow \begin{bmatrix} C_{1,1,1} \\ \vdots \\ C_{1,\Theta,K} \end{bmatrix},$$

其中 $r \times o = m$ 和 $r \times p = n$, 其中 r 为汇聚半径或窗口大小, 另外 C1 层上的响应为

$$C1_{i,j} = \max_{r \times r} \{S_{1,i,j}\}.$$

即给定窗口 $r \times r$, 对 S1 层上的响应通过不重叠的滑动窗处理, 对于每一个窗口内的 r^2 个元素取其最大

的元素,用此元素作为 C1 层上相应位置的响应,进而形成对输入图像具有局部的平移不变性描述.

S2 层:主要模拟视觉皮层 V4 区上的简单细胞的感受野特性,即带通性、平移不变性以及对角、色彩、纹理等中层复杂特征的偏好刺激等.具体的操作是先定义一簇新的滤波器,记为

$$f_j \in R^{a \times b \times \theta, k}, j = 1, 2, \dots, N,$$

其中 θ 为方向个数,与 S1 层的一致. k 为尺度个数且小于 S1 层的尺度个数 K . 该层上的响应为

$$S2: R^{o \times p \times \theta, K} \rightarrow R^{q \times r \times u}$$

$$\begin{bmatrix} C1_{1,1} \\ \vdots \\ C1_{\theta,K} \end{bmatrix} \rightarrow \begin{bmatrix} S2_1^j \\ \vdots \\ S2_u^j \end{bmatrix}.$$

注意 $o > a$ 和 $p > b$, 并且有

$$\begin{aligned} S2_i^j &= f_j \times C1_{\theta, i \sim i+k-1} \\ &= \sum_{\theta} \sum_t f_{j, \theta, t} \times C1_{\theta, t}, \end{aligned}$$

其中式子 $C1_{\theta, i \sim i+k-1} = [C1_{\theta, i}, C1_{\theta, i+1}, \dots, C1_{\theta, i+k-1}]^T$ 和 $f_j = [f_{j, \theta, 1}, f_{j, \theta, 2}, \dots, f_{j, \theta, k}]$ 且有 $f_{j, \theta, i} \in R^{a \times b}$; 另外 $i = 1, 2, \dots, u$ 且 $u = K - k, q = o - a$ 和 $r = p - b$.

C2 层:主要模拟 V4 区上的复杂细胞及 PIT/AIT 区上的简单细胞的感受野特性,即方向性、较好的平移不变特性和尺度不变特性等.具体的操作为

$$C2: R^{q \times r \times u} \rightarrow R$$

$$\begin{bmatrix} S2_1^j \\ \vdots \\ S2_u^j \end{bmatrix} \rightarrow F^j = \max_{\sigma} \{S2_{\sigma}^j\},$$

其中这里的 $\sigma = 1, 2, \dots, u$; 由于 S2 层中滤波器的选取与变量 j 相关且有 $j = 1, 2, \dots, N$. 所以最终 C2 层得到的特征为 $F = [F^1, F^2, \dots, F^N]^T \in R^{N \times 1}$.

沿着 HMAX 模型的结构,还有很多好的改进思路,除了 Theriault 的工作外, Serre 等人基于 HMAX 模型,在 S2 层上引入特征字典的学习过程,并进一步细化了每一层上的功能,使得改进后的模型针对各种复杂场景下的目标提取具有局部的平移和尺度不变性的特征.随着神经生理研究的发展,已经研究证实不仅在 V1 区上,几乎在整个腹侧视觉通路的每一视区皮层上,神经元都具有稀疏响应特性^[28-30]. 为了将这种稀疏性作为约束条件嵌入到 HMAX 模型中,2014 年清华大学的胡晓林博士等人^[11]提出了层次化稀疏模型,即 Sparse-HMAX,它仍采用 HMAX 模型的结构,但与其不同的是滤波器不再事先确定,而是通过稀疏编码或独立成分分

析从给定的数据集中学习得到的.

4 稀疏认知计算模型的研究进展

基于生物视觉稀疏认知机理的研究脉络,本节将综述稀疏认知计算模型的研究进展,包括稀疏编码模型、结构化稀疏模型和层次化稀疏模型这 3 种模型的结构、求解算法和发展现状,并阐释这 3 种模型的区别与联系.

4.1 稀疏编码模型

从 1996 年 Olshausen 和 Field 提出了稀疏编码之后,关于稀疏编码的研究与应用开启了继小波分析之后,信号与图像处理领域的又一次新浪潮. 下面,本节将从模型的结构、求解算法以及发展现状等 3 个方面来认识该模型.

4.1.1 模型的结构

基于 V1 区简单细胞的稀疏性学习与建模,通过 l_0 范数(伪范数)来约束编码系数,便得到了稀疏编码模型所求解的优化问题:

$$P_0: \min_{\alpha} \|x - D \cdot \alpha\|_2^2 + \lambda \|\alpha\|_0,$$

其中: x 表示信号或者信号的特征; D 表示过完备的字典; α 为编码系数; $\|\cdot\|_0$ 为向量的 l_0 范数,表示向量 α 中非零元素的个数.

由于 P_0 问题是一个非凸优化问题,并且求解是 NP-Hard 难的. 所以,斯坦福大学的 Jibshirani 教授和加州大学伯克利分校的 Breiman 教授几乎同时提出对编码系数施以 l_1 范数的正则约束,利用最小绝对收缩和选择算子(即 Lasso),促使求解出来的编码系数尽可能的稀疏,使得 P_0 优化问题放松为下面的 P_1 优化问题:

$$P_1: \min_{\alpha} \|x - D \cdot \alpha\|_2^2 + \lambda \|\alpha\|_1,$$

其中的 $\|\alpha\|_1 = \sum_k |\alpha_k|$, 即向量 α 中所有元素绝对值的和.

另外,2006 年 Candes 等人^[31]合作证明了在限制等距特性的条件下, l_0 范数优化问题 P_0 与以下 l_1 范数优化问题 P_1 具有相同的解. 之后,学者们研究了介于 P_0 与 P_1 问题之间的优化问题^[32-34]:

$$P_p: \min_{\alpha} \|x - D \cdot \alpha\|_2^2 + \lambda \|\alpha\|_p,$$

其中 $0 < p < 1$, $\|\alpha\|_p = (\sum_k |\alpha_k|^p)^{1/p}$.

4.1.2 模型的求解

针对 P_0 问题,1993 年 Mallat 和 Zhang 等人提出了一种贪婪算法,即匹配追踪算法; 针对 P_1 问题,

1995年Chen和Donoho等人提出了一种凸松弛算法,即基追踪算法.之后随着对这两个问题形式的不断研究,形成了一系列有意义的贪婪算法和凸松弛算法.其中代表性的贪婪算法有:正交匹配追踪、弱匹配追踪和阈值算法等;代表性的松弛算法有基追踪和迭代收缩算法等.有关这些算法的详细描述以及其稳定性讨论请参考文献[35-36].对于 P_p 问题,Gorodnitsky和Rao等人提出了FOCUSS(Focal Under Determined System Solver)算法来求解该非凸优化问题,所得到的解是对 P_0 问题最优解的一种逼近.另外,当 p 在开区间 $(0, 1)$ 上变化时,国内学者徐宗本等人验证了 $p=0.5$ 时解的最优性[37].

4.1.3 模型的发展现状

随着稀疏编码模型被广泛的研究与使用,模型的优化问题已经演变为合成与分析两种形式,其中合成形式指的是如 P_0 、 P_1 等优化问题,求解得到的是稀疏表示系数.该形式的稀疏模型的理论研究(包括模型的求解算法及稳定性分析,字典学习理论等)已经比较完善.

相比于合成稀疏模型,分析稀疏模型的研究相对比较年轻,模型的结构可以写为

$$\min_x \|y - \mathbf{M} \cdot x\|_2^2 + \lambda \|H \cdot x\|_0,$$

其中: y 为观测信号; \mathbf{M} 为观测矩阵; x 为待恢复的原始信号; H 为分析字典,它可以理解为合成稀疏模型中字典 D 的稀疏对偶[38], $H \cdot x$ 为 x 经过 H 变换之后得到的表示系数.关于分析稀疏模型的求解,Nam等人提出了贪婪分析匹配算法.有关模型的理论和应用可以参考2011年Nam和Davies等人的工作[39].

另外,为了研究这两种模型的异同,Elad等人[40]从理论上分析了这两种模型,得到“没有明确的表示这两种模型哪一个更好”的结论.但是2012年美国旧金山州立大学的李世东教授给出了这两种模型之间的联系,即在稀疏对偶的意义下任意一个合成稀疏模型都有一个等价的分析稀疏模型[38].

4.2 结构化稀疏模型

4.2.1 模型的结构

基于V1区复杂细胞的稀疏性学习与建模,通过引入局部的拓扑结构特性,即稀疏表示系数的结构信息,便可实现更为合理的特征学习模型.将表示系数的结构信息融入到稀疏编码模型中,便得到了结构化稀疏模型,其结构为

$$\min_{\alpha} \|x - D \cdot \alpha\|_2^2 + \lambda \Phi_G(\alpha),$$

其中: x 为信号; D 为字典;集合 G 为表示系数 α 的结构特性,这里结构特性可以是先验结构也可以是从样例集中学习出来的结构; $\Phi_G(\alpha)$ 为稀疏表示系数 α 在结构 G (如群集)上的正则约束.下面给出一种较为简单的群稀疏描述,即集合 G 可写为

$$\begin{cases} G = \{G_1, G_2, \dots, G_K\} \\ G_k = [\alpha^k(1), \alpha^k(2), \dots, \alpha^k(n_k)], k = 1, 2, \dots, K \end{cases},$$

其中群集 G 共有 K 个群, n_k 为群 G_k 的个数,每一个 $\alpha^k(i) \in \Lambda$,其中 $\Lambda = \{\alpha(1), \alpha(2), \dots, \alpha(m)\}$, m 为 α 的个数.那么 $\Phi_G(\alpha)$ 的正则约束,如 $l_{1,2}$ 范数,可写为

$$\Phi_G(\alpha) = \sum_{k=1}^K \|G_k\|_2 = \sum_{k=1}^K \left(\sum_{i=1}^{n_k} |\alpha^k(i)|^2 \right)^{1/2}.$$

更多有关数据的结构化学习可以参考文献[41-42].

4.2.2 模型的求解

稀疏编码模型利用 l_0 范数或 l_1 范数作为正则约束,得到的系数是非零个数层次上的稀疏,而现在的结构化稀疏模型利用表示系数的某种结构特性作为正则约束(如群稀疏),使得表示系数上升到了拓扑结构层次上的稀疏.针对不同信号的结构特性,所发展起来的结构化稀疏模型的求解算法也不尽相同,其中代表性的贪婪算法有:结构正交匹配追踪算法、块稀疏正交匹配追踪等;代表性的松弛算法有块稀疏匹配追踪,群Lasso算法等.有关结构化稀疏模型求解算法的总结可以参考文献[43-44].

4.2.3 模型的发展现状

随着结构化压缩感知理论的发展,结构化稀疏模型的研究与应用也引起了广泛的关注.其中有关结构化稀疏模型的理论分析可参考2009年到2011年法国学者Mairal和Jenatton等人的工作[45-46].同时,针对先验结构和后验结构这两种学习的模式,结构化稀疏模型也分为基于先验正则的结构化稀疏模型和基于后验正则的结构化稀疏模型,有关这方面工作的具体描述请参考文献[43].

另外,美国罗格斯大学的Huang和Zhang等人于2009年基于信息论编码法则给出不同特性数据在字典下的表示系数的5种结构形式,包括标准稀疏、群稀疏、层次稀疏、图稀疏和随机场稀疏;也推导出这5种结构形式的编码复杂度公式,结合实验分析得到:若信号对应的表示系数的编码复杂度越小,则利用编码复杂度作为非凸正则约束求解得到的表示系数的重构性能越好,并且能够反映出表示系数的结构特性[47].

为了有效地利用稀疏表示系数的结构信息,通常采用的是贝叶斯方法,因为该方法较容易嵌入结

构特性,这一点可参考武汉大学孙洪等人^[44]给出的基于贝叶斯模型下从稀疏到结构化稀疏的论文,关于结构化稀疏模型的更多最新研究进展,可以浏览 Jenatton 的个人主页,以及 Koray 和 Rob 等人^[48]基于结构化稀疏模型的不变量特征学习的工作。

4.3 层次化稀疏模型

4.3.1 模型的结构

基于腹侧视觉通路的稀疏性学习与建模,通过在 HMAX 模型的每一简单层上引入稀疏正则约束

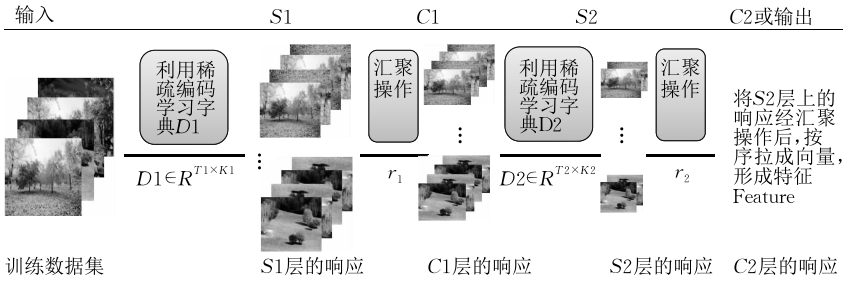


图 4 具有 4 层的层次化稀疏模型的训练结构

当训练完成后,利用训练得到的模型参数,即简单层响应前的字典;便可以实现对于测试图的特征提取。

4.3.2 模型的学习

本小节,结合图 4 给出层次化稀疏模型的具体学习方式,分为两部分:即训练与测试。

(1) 训练部分

① 训练数据. $(x_i, y_i), 1 \leq i \leq L$, 其中 x_i 为一幅图像,大小为 $N \times M$, L 幅图像的大小相同; y_i 为相应的输出(如类标等)。

② S1 层前的字典学习. $D1 \in R^{T1 \times K1}$, 其中 $K1$ 为字典中基原子的个数, $T1$ 为每个原子的维数并且有 $T1 = p \times p$ 成立. 利用训练数据,对每一幅图像 x_i , 采样 n 个图像块,每一块大小为 $p \times p$, 将每一图像块拉成一 $T1$ 维的向量;最后便可得到一新数据集 $X \in R^{T1 \times (L \times n)}$, 其中 $L1 = L \times n$; 利用此数据集,基于下面的优化问题:

$$\min_{D1, \Lambda} \|X - D1 \cdot \Lambda\|_F^2 + \lambda \sum_{i=1}^{L1} \|a_i\|_1$$

$$\text{s. t. } \|d1_i\|_2^2 \leq 1, i = 1, 2, \dots, K1$$

其中: a_i 为 Λ 的第 i 列; $d1_i$ 为 $D1$ 的第 i 列; λ 为正则因子; $\|\cdot\|_F$ 为 Frobenius 范数. 求解该问题便可以学习字典 $D1$.

③ S1 层上的响应. 利用得到的字典,通过取伪逆便可以得到相应的滤波器集,即 $H1 = \text{pinv}(D1) \in R^{K1 \times T1}$ 其中 $H1$ 的每一行为一滤波器,记为 h_i^1 , 上角

并利用训练数据结合稀疏编码来训练该层响应的字典,如图 4 所示,给出了层次化稀疏模型的训练结构简图。

图 4 中 S1 层与 S2 层响应前的字典学习主要是为了刻画稀疏性的学习,而 C1 层与 C2 层响应前的汇聚操作是为了获取对目标的局部平移不变性的描述;其中 r_1 和 r_2 为汇聚半径(预先给定的参数), $D1$ 与 $D2$ 为需要训练的字典. 特别需要注意的是层次化稀疏模型可以是多层的结构,不局限于 4 层。

标 1 和下角标 i 表示来自于 $H1$ 的第 i 行及 $i = 1, 2, \dots, K1$. 对于一幅输入图像 x_k , 利用滤波器集 $H1$, 做卷积便可得到 S1 层上的响应. 即如下操作,先将每一滤波器

$$\bar{h}_i = \text{reshape}(h_i^1, T1, p, p) \in R^{p \times p},$$

这里的 reshape 为 Matlab 里的函数,即将长度为 $T1$ 的行向量 h_i^1 变为 $p \times p$ 的矩阵 \bar{h}_i , 然后与 x_k 卷积得

$$y_{k,i} = x_k \times \bar{h}_i \in R^{(N-p+1) \times (M-p+1)}.$$

由于共有 $K1$ 个滤波器,所以有

$$Y_k = \{y_{k,i}\}_{i=1}^{K1}.$$

注意这只是第 k 幅输入图像的 S1 层上的响应,最终 L 幅训练图像的 S1 层响应为

$$Y = \{Y_k\}_{k=1}^L.$$

④ C1 层上的响应. 利用 S1 层上响应 Y , 给定 C1 层响应前汇聚操作的半径为 r_1 , 即将 Y 中的每幅图,如 $y_{k,i}$, 通过利用窗口大小为 $r_1 \times r_1$ 无重叠地对 $y_{k,i}$ 进行分割,每一分割区域中的 $r_1 \times r_1$ 个元素的最大值来表示该区域,即最大汇聚操作,可以得到

$$\bar{y}_{k,i} \in R^{\frac{(N-p+1)}{r_1} \times \frac{(M-p+1)}{r_1}}.$$

实验中 r_1 尽可能地选取能被 $N-p+1$ 和 $M-p+1$ 都整除的数. 最终得到 C1 层上的响应为

$$\bar{Y} = \{\bar{Y}_k\}_{k=1}^L = \{\{\bar{y}_{k,i}\}_{i=1}^{K1}\}_{k=1}^L.$$

⑤ S2 层前的字典学习. $D2 \in R^{T2 \times K2}$, 其中 $K2$ 为字典中基原子的个数, $T2$ 为每个原子的维数并且有 $T2 = q \times q \times K1$ 成立. 根据 C1 层上的响应 \bar{Y} , 对

于 \bar{Y}_k 而言, 采样 m 个图像“块”, 每一“块”的大小为 $q \times q \times K1$, 简图如图 5.

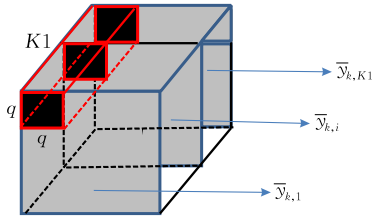


图 5 \bar{Y}_k 中的图像“块”采样

即有 $K1$ 个大小为 $q \times q$ 的小块, 将每一图像“块”拉成一 $T2$ 维向量; 对于 \bar{Y}_k 共采样 m 个图像“块”, 因而有 m 个 $T2$ 维向量. 最终对于 \bar{Y} 而言, 每一 \bar{Y}_k 都采样 m 个图像“块”, 所以最终可以形成一新的数据集, $\bar{X} \in R^{T2 \times (L2)}$, 其中 $L2 = L \times m$; 利用此数据集, 基于下面的优化问题

$$\begin{aligned} \min_{D2, \Omega} \quad & \|\bar{X} - D2 \cdot \Omega\|_F^2 + \gamma \sum_{i=1}^{L2} \|\omega_i\|_1 \\ \text{s. t.} \quad & \|d2_i\|_0 \leq 1, \quad i = 1, 2, \dots, K2 \end{aligned}$$

其中: ω_i 为 Ω 的第 i 列; $d2_i$ 为 $D2$ 的第 i 列; γ 为正则因子. 求解该问题便可以学习字典 $D2$.

⑥ $S2$ 层上的响应. 与 $S1$ 层上的操作类似, 先将字典取伪逆得到滤波器集, 即 $H2 = \text{pinv}(D2) \in R^{K2 \times T2}$; 然后对于 $H2$ 的每一行, 如第 i 行 h_i^2 , $i = 1, 2, \dots, K2$, 为一滤波器; 对于 $C1$ 层上的响应而言, “每一幅” \bar{Y}_k 作为 $C1$ 层上的输出, 与 $H2$ 的每一行来进行卷积便得到 $S2$ 层上的响应. 操作如下, 先将每一滤波器

$$\hat{h}_i = \text{reshape}(h_i^2, T2, q, q, K1) \in R^{\rho \times \rho \times K1},$$

这里的 reshape 不再是 Matlab 中的函数, 但其含义类似, 是将长度为 $T2$ 的行向量 h_i^2 变为 $K1$ 个 $q \times q$ 大小的滤波器; 然后与 \bar{Y}_k 卷积得到

$$\begin{aligned} z_{k,i} &= \bar{Y}_k \times \hat{h}_i \\ &= \sum_{t=1}^{K1} \bar{y}_{k,t} \times \hat{h}_{i,t} \in R^{\left(\frac{(N-p+1)}{r_1} - q + 1\right) \times \left(\frac{(M-p+1)}{r_1} - q + 1\right)}, \end{aligned}$$

其中 $\hat{h}_{i,t}$ 为 \hat{h}_i 的第 t 个 $q \times q$ 大小的滤波器, 其中 t 为 $1, 2, \dots, K1$. 随着 $i = 1, 2, \dots, K2$ 的变化, 可以得到 \bar{Y}_k 在 $S2$ 层上的响应为

$$Z_k = \{z_{k,i}\}_{i=1}^{K2}.$$

最终, 当 $k = 1, 2, \dots, L$ 变化时, 即所有的训练图像得到的 $S2$ 层上的响应为

$$Z = \{Z_k\}_{k=1}^L.$$

⑦ $C2$ 层上的响应. 利用 $S2$ 层上的响应 Z , 给定 $C2$ 层响应前汇聚操作的半径为 r_2 , 即将 Z 中的每一幅图, 如 $z_{k,i}$, 通过利用窗口大小为 $r_2 \times r_2$ 无重叠

地对 $z_{k,i}$ 进行分割, 通过最大汇聚的操作得到

$$\bar{z}_{k,i} \in R^{\left(\frac{(N-p+1)}{r_1} - q + 1\right) \times \left(\frac{(M-p+1)}{r_2} - q + 1\right)},$$

得到 $C2$ 层上的响应为

$$\bar{Z} = \{\bar{Z}_k\}_{k=1}^L = \{\{z_{k,i}\}_{i=1}^{K2}\}_{k=1}^L.$$

注意这一步需要利用 \bar{Z} 来形成特征, 方法是将其按序拉成一列, 即

$$\begin{cases} \text{Feature}(1) = [z_{1,1}, z_{1,2}, \dots, z_{1,K2}]^T \\ \text{Feature}(2) = [z_{2,1}, z_{2,2}, \dots, z_{2,K2}]^T \\ \vdots \\ \text{Feature}(L) = [z_{L,1}, z_{L,2}, \dots, z_{L,K2}]^T \end{cases}$$

其中 $z_{k,i}$ 为一个行向量, 它是矩阵 $z_{k,i}$ 按行排列级联而成的. $\text{Feature}(i)$ 表示得到的第 i 幅训练图像的特征, 并且有 $\text{Feature}(i) \in R^{P \times 1}$, $i = 1, 2, \dots, L$. 其中

$$\begin{aligned} P &= \left(\frac{(N-p+1)}{r_1} - q + 1\right) \times \left(\frac{(M-p+1)}{r_2} - q + 1\right) \times K2. \end{aligned}$$

(2) 测试部分

当训练过程中的字典 $D1$ 与 $D2$ 学习好之后, 汇聚半径 r_1 和 r_2 已给定, 那么对于一幅测试图像, 依据上面的分析过程, 便可以学习得到其特征. 可以看出该特征的学习是一种无监督的方式. 进一步, 利用训练图像得到的特征和类标信息可以学习一个分类器, 进而利用测试图像的特征对测试图像实现分类任务.

4.3.3 模型的发展现状

基于腹侧视觉通路的稀疏性学习与建模的研究, 层次化稀疏模型已成为稀疏认知计算模型的又一研究热点. 从 1999 年 Riesenhuber 和 Poggio 等人提出的 HMAX 模型到 Serre 等人将特征字典的学习引入到 HMAX 模型中, 再到 2014 年 Hu 等人^[11]将稀疏性作为约束条件嵌入到 HMAX 模型的近 15 年时间里, HMAX 的算法复杂度在不断改进, 逐渐接近对生物视觉皮层腹侧通路机理的正确理解.

在这里需要强调的是, 层次化稀疏模型与以往的深度神经网络有所不同, 图 6 简单地给出了层次化稀疏模型与深度卷积神经网络^[49]的对比.

4.4 3 种模型之间的区别与联系

首先, 从生物视觉的稀疏认知机理来看, 稀疏编码模型, 结构化稀疏模型和层次化稀疏模型分别是为了刻画 V1 区简单细胞感受野的特性, V1 区复杂细胞感

	层次化稀疏模型	深度卷积网络
模型结构	输入 $\rightarrow S_1 \rightarrow C_1 \rightarrow S_2 \rightarrow C_2 \rightarrow \dots \rightarrow$ 输出	输入 $\rightarrow C_1 \rightarrow S_1 \rightarrow C_2 \rightarrow S_2 \rightarrow \dots \rightarrow$ 输出
S,C含义	S \rightarrow 简单细胞单元 C \rightarrow 复杂细胞单元	C \rightarrow 卷积Convolution S \rightarrow 次采样及汇聚操作
结构模块	{ 稀疏字典学习层 汇聚操作层	{ 卷积层或滤波层 次采样及汇聚操作
学习得到的字典与滤波器特性	稀疏字典学习层,学习到的字典有局部化、方向性和带通特性;最主要的是该层能够刻画输入刺激的稀疏特性.汇聚层主要是为了获取局部平移不变的特性	卷积层或滤波层,学习到的滤波器有两个特性,即局部化和权值共享特性;次采样与汇聚层主要是为了获取局部平移不变的特性
优化策略	优化字典是通过在每一简单细胞单元层前进行,利用K-SVD算法便可得到.(局部单元层的优化问题)	优化滤波器及偏置是通过在建立输入与输出的模型后,构建损失函数及准则或正则约束来建立优化目标函数,进而利用梯度下降法来求解.(全部单元层的优化问题)

图 6 层次化稀疏模型与深度卷积网络的对比

受野的特性和视觉皮层腹侧通路的机理而提出的。

其次,从稀疏认知计算模型的结构来分析,结构化稀疏模型是在稀疏编码的基础上,通过将表示系数的结构信息融入到稀疏约束项中而得到的;层次化稀疏模型也是在稀疏编码的基础上,通过在每一简单层上利用稀疏编码来学习字典,层次化稀疏模型中每一复杂单元层上的(局部最大)汇聚操作相当于聚类,是为了更好地描述简单单元层上响应(即稀疏表示系数)的结构特性;另外从模型的结构来看,层次化稀疏模型可视是为结构化稀疏模型通过逐层学习机制得到的,虽然它与结构化稀疏模型在获取结构的操作上有所不同,但都是通过加入表示系数的结构信息使得模型的最终响应具有局部的变换不变性。

最后,从应用的角度来看,通过稀疏编码或结构化稀疏模型求解得到的稀疏表示系数可以成功地应用在信号处理中的重构、压缩、修复与去噪等任务中,但将稀疏表示系数作为一种特征来实现目标分

类却不那么合理奏效^[50],原因是因为过完备字典的条件数往往比较大,导致稀疏表示系数的微小变化并不能对应着目标的微小变化;而通过层次化稀疏模型求解得到的输出响应却可以作为一种有效的特征来实现目标的分类,原因是因为该模型采用了稀疏编码与汇聚操作相结合的处理方式,使得模型得到的输出可以较好地刻画输入信号的特性。

5 基于稀疏认知计算模型的目标识别

5.1 稀疏认知学习、计算与识别范式的脉络结构

当“稀疏”一词被提出以后,各种与之有关的理论、模型和算法等如雨后春笋般涌现出来,如稀疏表示理论^[51]、稀疏贝叶斯学习理论^[52]、稀疏编码模型、稀疏主成分分析^[53]和稀疏优化算法^[54-55]等等.为了更好地理解本文提出的稀疏认知学习、计算与识别这一范式,图7将通过简图呈现这一范式的脉络结构。

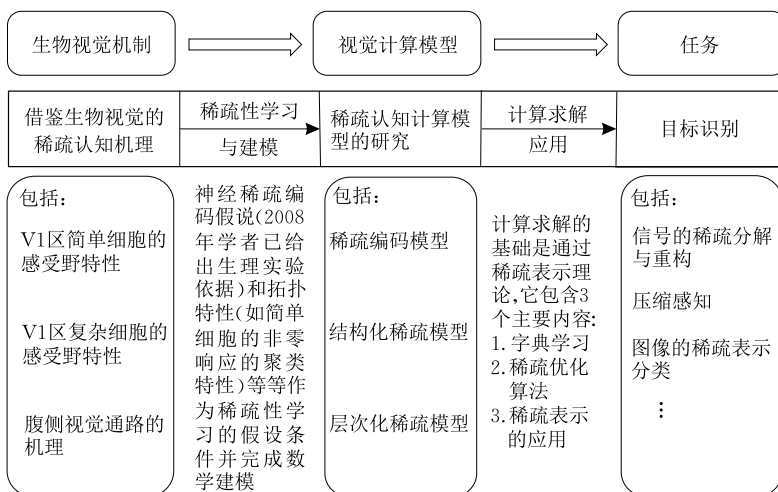


图 7 稀疏认知学习、计算与识别范式的脉络结构

注意:在本文的第2节 V1 区简单细胞的稀疏性学习与建模中,得到的优化问题由两部分构成,即信息保真项和稀疏约束项.关于稀疏约束项,我们采用的是以 l_0 或 l_1 范数来描述的,进而通过稀疏表示理论中的算法进行求解,需要强调的是关于稀疏约束项还有不同的描述方式,如稀疏自编码器中利用的 KL 距离^[56].

另外需要注意的是稀疏表示与稀疏编码之间的关系.稀疏表示是指,信号可以用一个过完备的字典进行线性表示,而该表示对应着欠定的线性方程组(即表示系数不唯一);为使表示系数被唯一求解,从而约束表示系数是稀疏的(即仅有少数的表示系数为非零值).根据神经稀疏编码假说,即神经生理发现,V1 区上的细胞大致有 5000 万个,而外侧膝状体上的细胞只有 100 万个左右,因而 V1 区细胞编码表达空间的维数大于输入空间(即外侧膝状体细胞编码表达空间)的维数,所以亦为一个欠定的线性方程组,为有效地描述神经元对外界刺激的响应特性,故假设 V1 区上的神经元采取的是稀疏编码的原则(该假设已证实).从得到的问题来看,稀疏表示与稀疏编码是一致的;从模型的结构来分析,它们都是由两部分构成,即信息保真项和稀疏约束项;不同的是二者刻画稀疏约束的方式不同,稀疏表示是通过 l_0 或 l_1 范数,而稀疏编码是通过这些函数 $-e^{-x^2}$, $|x|$ 和 $\log(1+x^2)$ 等形式.需要注意的是本文所述的稀疏编码模型是以稀疏表示理论(即以 l_0 或 l_1 范数作为稀疏约束条件)为基础而展开的.

5.2 稀疏认知计算模型的应用概述

5.2.1 稀疏编码模型的应用

从稀疏认知计算模型的发展来看,有关稀疏编码模型的研究已经历了近 18 年,其相应的应用也遍及信号与图像处理的各种任务,如 Donoho 和 Candès 等人^[57-59]基于稀疏编码模型提出的压缩感知理论,已成功应用在医学信号、自然信号等压缩处理任务中,有关压缩感知的综述可以参考文献^[60-64];以及 Elad 等人^[65-73]将稀疏编码模型应用在自然图像的去模糊、降噪、压缩、分割、修复和超分辨等各种任务上,并也取得了较好的研究成果.

另外,该模型结构下的一些代表性工作有 2009 年 Wright 等人^[74-77]提出的稀疏表示分类模型(SRC),已成为人脸图像分类的经典算法,此后对于 SRC 模型的研究与改进(如核化稀疏表示模型)也成为了自然图像分类任务中的一个研究热点;和 2010 年美国亚利桑那州州立大学的 Zhang 和 Li 等

人^[9]针对人脸图像分类的任务,提出了判别稀疏模型,该模型仍基于稀疏编码模型,充分地利用了训练样例集的分类信息,使得到的模型一方面具有稀疏表示能力,另一方面也具有判别分类的能力.

5.2.2 结构化稀疏模型的应用

关于结构化稀疏模型的研究也发展了 10 余年,其应用性的代表工作有:2009 年 Huang 和 Zhang 等人^[47]提出的基于信息论编码法则约束的结构化稀疏模型;2012 年 Han 和 Wu 等人^[78]提出的用于图像标注具有输入输出约束的结构化稀疏模型;以及同一年 Pontil 和 Mourao-Miranda 等人^[79]利用大脑的功能性核磁共振成像数据来解码大脑的结构化稀疏模式,并对结构化稀疏模型进行了稳定性分析.

另外基于对 SRC 模型的分析,2012 年 Jia 等人^[80]又提出了具有鲁棒性的结构化稀疏表示分类器,他们认为之前利用测试图像与其逼近图像(通过字典与稀疏表示系数相乘得到的)之间的差图像,并通过 l_1 范数分别作用在表示系数和差图像上来促使稀疏性,其中作用在差图像上的方式是不合理的,因为差图像中的每一个像素不能被独立对待,而应被看作是具有连续和空域局部化的结构特性,所以引入了差图像上的结构作为稀疏正则后,便改善了 SRC 的分类性能;2012 年 Zhang 和 Huang 等人^[81]将结构化稀疏模型应用在自然图像去模糊的任务上取得了较好的结果;以及 2014 年 Asaei 等人^[82]基于混响室中语音的结构特性进行假设并建模,利用结构化稀疏模型应用到语音分离的任务中,取得了较好的分离结果.

5.2.3 层次化稀疏模型的应用

相比稀疏编码模型和结构化稀疏模型,层次化稀疏模型的形成与发展的时间较晚,但是它能够获取输入图像的一些高层复杂或深层特征,进而实现复杂场景下的目标识别.其代表性的工作可以参考 2014 年 Hu 等人将该模型用于自然图像分类.由于层次化稀疏模型是对 HMAX 模型的改进,进而也是一种层次化目标识别模型,那么许多基于 HMAX 模型的应用,如图像的分类、分割、轮廓提取、人脸表情识别和语音识别等^[83-87],也可以利用层次化稀疏模型来解决.

其中基于 HMAX 模型的语音识别^[83],是 2014 年 Roos 和 Wolmetz 等人的工作,根据听觉神经科学的最新研究显示听觉与视觉皮层分享相似的结构与功能组织,基于此相似性,他们将 HMAX 模型用到了语音识别中的词识别任务上. 通过从 TIMIT

语音库中记录一些语音样本(训练样本),利用听觉外周的计算模型将语音样本转化为时频谱图,进一步将这些谱图分成 750 ms 的帧,并将这些帧输入到 HMAX 模型中得到相应的特征描述,进而利用这些特征训练 SVM 分类器(一对多方式)去辨识包含目标词的帧;而后对于一套测试句子,获取其时频谱图,采用滑帧的方法,得到该句子中的每一帧,进而利用 HMAX 模型获取其特征描述,并利用训练好的 SVM 分类器来实现是否该测试句中有目标词被提出.类似的,可以将 HMAX 模型利用层次化稀疏模型来替代,由此而带来的好处是获取帧的特征更具稀疏性.

5.3 稀疏认知计算模型目标识别的示例

基于稀疏认知学习、计算和识别范式,图 8 简要地给出该范式应用于目标识别的框架.

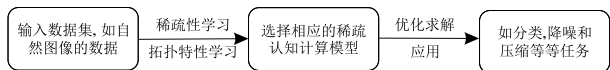


图 8 稀疏认知学习、计算与识别范式的目标识别框架

利用此框架,下面通过举例来说明这种范式已获得的成果.

5.3.1 基于稀疏表示的人脸识别

人脸识别是基于人的脸部特征信息进行身份识别的一种生物识别技术.2009 年 Wright 等人提出了稀疏表示分类(SRC)模型,已成功地应用在人脸识别上并取得了较好的分类结果.下面,利用 Wright 等人文中的一个实验来详述基于稀疏表示的人脸识别.首先,给出 SRC 的算法描述,

算法 1. 稀疏表示分类(SRC).

输入:字典 $D=[D_1, D_2, \dots, D_K] \in R^{m \times n}$, 其中 D_k 为第 k 类的子字典,即从第 k 类的训练样例集中选取一定数量的样例,按列级联而成;和一测试样例 $x \in R^m$

输出:测试样例 x 的类标

1. 对 D 中的每一列进行 l_2 范数归一化操作得 \bar{D} .
2. 求解优化问题,

$$\hat{\alpha} = \arg \min_{\alpha} \|\alpha\|_1 \quad \text{subject to} \quad \|x - \bar{D}\alpha\|_2 \leq \epsilon.$$

3. 计算残差, $r_k(x) = \|x - \bar{D}_k \delta_k(\hat{\alpha})\|_2^2, k=1, 2, \dots, K$.

其中 $\delta_k(\hat{\alpha})$ 为向量 $\hat{\alpha} \in R^{n \times 1}$ 的第 k 类所对应的表示系数.

4. 识别类标, $Label(x) = \arg \min_k \{r_k(x)\}$.

其次,利用 Yale B 数据库,即包含 38 个类,共计 2414 幅图像,其中每一类由一个对象在各种光照条件获取的图像构成;每一幅图像的大小为 192×168 . 通过从每一类中随机选取一半的图像构成训练样例集,即有 1207 幅图像,其余作为测试样例集.然后再通过图像子采样对数据库中的每一幅图像进行

处理得到大小为 12×10 的子图像,将该子图像按行拉成 120 维的向量,该向量视为原图像的“特征”,对应着原来的训练数据集和测试数据集,便可以得到训练数据特征集和测试数据特征集.

最后,利用全部的训练数据特征集,级联形成字典 $D \in R^{120 \times 1207}$; 对于一个测试数据特征集,便可以利用 SRC 算法识别其类标.当 1207 个测试数据特征集完成类标预测后,统计结果得到 92.1% 的正确率.

5.3.2 基于层次化稀疏模型的目标识别

基于层次化稀疏模型的学习方式,即可学习输入图像的深层特征,进而利用训练图像的深层特征来训练多类线性 SVM 分类器,最终实现测试图像的分类任务.下面,利用胡晓林等文中的一个实验来理解层次化稀疏模型的目标识别任务.

首先,有关 4 层的层次化稀疏模型的参数设置如下, S_1 层前的 $D_1 \in R^{64 \times 8}$, 求解 D_1 时的优化问题中,新数据集 $X \in R^{64 \times 100000}$ 且 $\lambda = 0.15$, S_2 层前的 $D_2 \in R^{16 \times 1024}$, 新数据集 $\bar{X} \in R^{16 \times 200000}$ 且 $\gamma = 0.15$, 另外 C_1 层前的汇聚半径 $r_1 = 6$, C_2 层不再直接利用汇聚操作,而是先对 S_2 层的响应进行网格分辨率为 1, 2, 4 的三级空域金字塔处理^[88], 然后在每一个网格内进行最大汇聚操作.最终得到每一幅输入图像的深层特征的维数为 $1024 \times 21 = 21504$.

其次,利用 Caltech-101 数据库,即有 102 类,共计 9144 幅图像,每一类的图像个数从 31~800 不等,大部分的图像是中等分辨率.从每一类中随机的选取 15 幅图像构成训练样例,其余的作为测试样例.

最后,利用训练样例得到的深层特征训练 102 类的线性 SVM 分类器;对于测试样例中的每一幅图像便可以根据 4 层的层次化稀疏模型提取特征,之后利用训练好的 SVM 分类器识别其类标.最终得到的准确率为 66.45 ± 0.52 . 另外,若每类中选取 30 幅图像构成训练样例时,得到准确率为 73.67 ± 1.23 .

6 存在的问题及未来研究的方向

目前,虽然关于稀疏认知学习、计算与识别范式的理论研究与应用很多,也取得了较好的研究成果,但在该范式的学习过程中仍有许多问题尚待解决和完善.下面,基于该范式,利用假设、建模、求解和反馈调整这一数学建模的流程,给出该范式存在的一些问题,并指出可进一步研究的方向,即从生物视觉的神经稀疏编码假设出发(或信号处理中关于信号

的稀疏性假设), 建立描述该假设的视觉计算模型(或信号的稀疏表示模型), 通过设计的优化算法求解, 最后利用模型的结果来反馈调整假设条件, 修改模型直至假设条件能够合理地反映输入输出之间的关系。

6.1 存在的问题

6.1.1 假设中的问题

从假设的角度来分析, 稀疏编码模型能够模仿生物视觉皮层中神经元细胞的稀疏特性, 但从信号表示的角度来看, 该模型是依据信号与字典的基原子之间呈线性关系的假设而建立的; 若信号与字典的基原子之间是通过高度非线性关系来描述的, 那么再使用基于线性假设的稀疏编码模型就有可能在具体的应用任务中失效。

6.1.2 建模中的问题

从建模的角度来分析, 稀疏认知计算模型, 特别是针对稀疏编码模型和结构化稀疏模型, 它的实质可以通过图 9 的框架结构来理解。

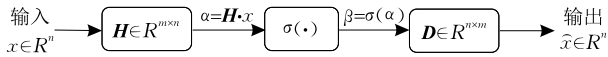


图 9 稀疏认知计算模型的框架结构

其中: x 为输入信号; \hat{x} 为输出信号, 是 x 的逼近; 矩阵 \mathbf{H} 和 \mathbf{D} 分别为分析字典和合成字典且 $m > n$; $\sigma(\cdot)$ 为非线性函数, 如 $\sigma(t) = |t|$, $\sigma(t) = \max(0, t - \theta)$ (其中 θ 为一阈值) 等。根据输入与输出之间的关系, 可得到

$$\hat{x} = \mathbf{D} \cdot \sigma(\mathbf{H} \cdot x),$$

进而定义损失函数

$$L(x; \mathbf{H}, \mathbf{D}) = \|x - \hat{x}\|_2^2 = \|x - \mathbf{D} \cdot \sigma(\mathbf{H} \cdot x)\|_2^2.$$

(1) 假设 \mathbf{D} 已知, \mathbf{H} 是未知的, 那么可以不考虑 \mathbf{H} 的处理, 直接得到损失函数

$$L(x; \beta) = \|x - \hat{x}\|_2^2 = \|x - \mathbf{D} \cdot \beta\|_2^2.$$

但是由于字典 \mathbf{D} 是过完备的, 故求解得到的 β 是不唯一的, 所以基于 β 稀疏性正则(如 β 的 l_1 范数)的优化问题便是稀疏编码模型, 即

$$\min_{\beta} \|x - \mathbf{D} \cdot \beta\|_2^2 + \lambda \|\beta\|_1.$$

若 $\sigma(t) = |t|$, 则得到非负稀疏编码模型, 即

$$\min_{\beta} \|x - \mathbf{D} \cdot \beta\|_2^2 + \lambda \|\beta\|_1$$

$$\text{subject to } \beta \geq 0$$

另外, 基于 β 的结构性稀疏约束, 便可以得到结构化稀疏模型。

(2) 假设 \mathbf{H} 已知, \mathbf{D} 是未知的, 并且 $\sigma(t) = |t|$, 则可以得到分析形式的稀疏编码模型

$$\min_{\hat{x}} \|x - \hat{x}\|_2^2 + \lambda \|\mathbf{H} \cdot \hat{x}\|_1.$$

注意, 由于 $\sigma(t) = |t|$, 所以之前 β 的稀疏性可以等价于 α 的稀疏性, 即 $\|\beta\|_1 = \|\alpha\|_1 = \|\mathbf{H} \cdot x\|_1$. 更为普遍的分析稀疏编码模型为

$$\min_{\hat{x}} \|y - \mathbf{M} \cdot \hat{x}\|_2^2 + \lambda \|\mathbf{H} \cdot \hat{x}\|_1,$$

其中: $y \in R^r$ 为观测信号; $\mathbf{M} \in R^{r \times n}$ 为观测矩阵且具有 $r \leq n$, \hat{x} 为待恢复的原始信号。

(3) 假设 \mathbf{H} 与 \mathbf{D} 都是未知的; $\sigma(\cdot)$ 函数给定, 该框架结构可以认为是一种稀疏自编码器^[89], 即通过训练样例学习得到 \mathbf{H} 与 \mathbf{D} , 进而对一测试信号 x , 其编码过程为 $\beta = \sigma(\mathbf{H} \cdot x)$; 解码过程为 $\hat{x} = \mathbf{D} \cdot \beta$.

那么结合上述的分析, 总结问题如下: 一是能否通过该框架结构的逐层学习, 来学习信号的深层特征, 即如何将稀疏编码模型推广至深度学习^[90]的模式? 二是本文 4.4 节曾提到通过求解稀疏编码模型得到的稀疏表示系数并不能作为一种特征来描述输入信号, 但该框架结构中的非线性函数 $\sigma(\cdot)$ 能否改善这一点, 从而在得到“特征” β 后嵌入线性分类器实现分类任务?

6.1.3 反馈调整中的问题

从反馈调整的角度来分析, 层次化稀疏模型能够大致地模拟腹侧视觉通路的机理, 它也为计算机视觉的发展提供了一种新的思路, 但与 HAMX 模型一样, 它仍是一种由简单单元层与复杂单元层交替组成的前馈等级结构, 并没有考虑视觉皮层间存在着反馈链接^[91-92]。

6.2 进一步研究的方向

针对假设中的问题, 一种可行性的策略便是引入核技巧(例如线性核、多项式核、高斯核^[93]等), 可以将信号和字典映射到高维变换空间上, 使其呈现线性关系, 再利用稀疏编码模型来进行求解. 这种基于核的稀疏编码模型是一个值得进一步研究和完善的方向。

对于建模中的问题, 利用逐层学习^[94-95]的方式, 将稀疏编码模型推广至深度学习的模式, 是一个全新的值得研究的方向. 另外基于稀疏认知计算模型的框架结构, 通过嵌入分类器(如线性分类器)来快速地实现分类任务将有着广泛的应用前景, 即如下的结构(见图 10)是值得深入研究的一个方向。

对于反馈调整中的问题, 为了考虑皮层间的反馈链接, 即通过先验知识或上下文信息的帮助, 需将“自底向上(前馈)”与“自顶向下(反馈)”相结合来理解生物视觉皮层腹侧通路的机理, 以实现复杂场景

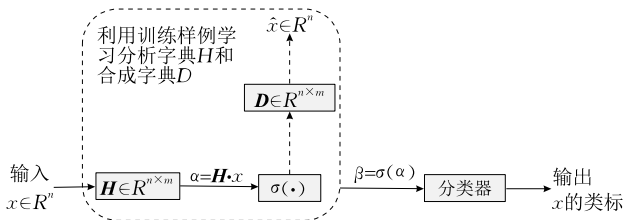


图 10 基于稀疏认知计算模型的分类结构

下目标的有效识别. 因此, 基于生物视觉皮层的研究, 构建高效性、高性能的层次目标识别模型仍是今后值得探索的热点方向.

另外, 针对稀疏认知计算模型中单一尺度上的基原子不能有效地表征具有奇异性的信号这个问题, 可借鉴小波分析, 多尺度几何分析等方法^[96-99], 构造多尺度字典来刻画信号的奇异特性, 这种多尺度字典的构造方法也是进一步需要研究的方向.

7 结束语

为了使得稀疏认知学习、计算与识别这一新范式具备学习能力, 高容量的表达能力, 快速推断能力和多任务信息共享能力, 借鉴生物视觉皮层的认知机理已成为一种必然的趋势^[26,100-101]. 目前, 越来越多的神经生理实验已证实, 神经元的稀疏编码特性几乎贯穿于整个腹侧视觉通路的每一皮层, 不只在V1区上. 本文正是以生物视觉稀疏认知机理的研究进展为依据, 通过对生物视觉稀疏认知机理的学习及建模, 详细地评述了稀疏编码模型、结构化稀疏模型和层次化稀疏模型的研究进展, 并阐释这3种模型之间的区别与联系; 以及概述稀疏认知计算模型目标识别应用的研究状况, 并举例给出了稀疏认知计算模型的具体成果; 最后总结了这一新范式存在的一些问题及可进一步研究的方向, 以期引起更多的研究同行对这一具有交叉性和发展性方向的关注.

致 谢 感谢陕西省智能感知与大数据协同创新中心、教育部创新团队和首批陕西省科技创新团队计划以及国家“九七三”重点基础研究发展规划和国家自然科学基金委重大研究计划的支持. 感谢计算机学报编辑部和评委老师的宝贵意见!

参 考 文 献

- [1] Hubel D H, Wiesel T N. Receptive fields of single neurones in the cat's striate cortex. *Journal of Physiology*, 1959, 148(3): 574-591
- [2] Mallat S, Zhang Z. Matching pursuit with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 1993, 41(2): 3397-3415
- [3] Willshaw D J, Buneman O P, Higgins H C L. Non-holographic associative memory. *Nature*, 1969, 222(5): 960-962
- [4] Barlow H B. Single units and sensation: A neuron doctrine for perceptual psychology. *Perception*, 1972, 1(4): 371-394
- [5] Olshausen B A, Field D J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 1996, 381(11): 607-609
- [6] Rodolphe J, Guillaume O, Bach F. Structured sparse principal component analysis. *Journal of Machine Learning Research*, 2010, 9(2): 366-373
- [7] Rodolphe J, Audibert J Y, Bach F. Structured variable selection with sparsity-inducing norms. *Journal of Machine Learning Research*, 2011, 12(1): 2777-2824
- [8] Duarte M F, Eldar Y C. Structured compressed sensing: From theory to application. *IEEE Transaction on Signal Processing*, 2011, 59(9): 4055-4085
- [9] Zhang Q, Li B X. Discriminative K-SVD for dictionary learning in face recognition//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. San Francisco, USA, 2010: 2691-2698
- [10] Serre T, Wolf L, Bileschi S, et al. Robust object recognition with cortex-like mechanisms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, 29(3): 411-426
- [11] Hu X L, Zhang J W, Li J M, Zhang B. Sparsity-regularized HMAX for visual recognition. *PLoS ONE*, 2014, 9(1): 1-12
- [12] Riesenhuber M, Poggio T. Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 1999, 2(10): 1019-1025
- [13] Zhao Song-Nian, Yao Li, Jin Zhen, et al. The whole video features of sparse representation in the human primary visual cortex: Evidence of brain functional imaging. *Chinese Science Bulletin*, 2008, 53(11): 1296-1304 (in Chinese)
(赵松年, 姚力, 金真等. 视像整体特征在人类初级视觉皮层上的稀疏表象: 脑功能成像的证据. *科学通报*, 2008, 53(11): 1296-1304)
- [14] Houweling A R, Brecht M. Behavioural report of single neuron stimulation in somatosensory cortex. *Nature*, 2008, 451(7174): 65-68
- [15] Zhou Yang, Wang Jian. The functional organization of the visual cortex and progress from fMRI. *Progress in Modern Biomedicine*, 2006, 6(9): 79-81 (in Chinese)
(周扬, 王健. 视觉皮层分区及其fMRI研究进展. *现代生物医学进展*, 2006, 6(9): 79-81)
- [16] Yao Xing-Zhong, Lu Tong-Wei, Hu Han-Ping. Object recognition models based on primate visual cortices: A review. *Pattern Recognition and Artificial Intelligence*, 2009, 3(4): 581-588 (in Chinese)
(姚行中, 鲁统伟, 胡汉平. 基于灵长类视觉皮层的目标识别模型综述. *模式识别与人工智能*, 2009, 3(4): 581-588)

- [17] Li Yi-Ping, Zhao Dan-Dan, Ma Lin, Liu Yi. Research advances on receptive field. *West China Medical Journal*, 2008, 23(3): 640-641(in Chinese)
(李奕萍, 赵丹丹, 马麟, 刘谊. 感受野研究的现状和进展. *华西医学*, 2008, 23(3): 640-641)
- [18] Shang Li. A Study of Sparse Coding Algorithm and Their Application [Ph. D. dissertation]. University of Science of Technology of China, Hefei, 2006(in Chinese)
(尚丽. 稀疏编码算法及其应用研究[博士学位论文]. 中国科学技术大学, 合肥, 2006)
- [19] Jiao Li-Cheng. *Neural Network Computation*. Xi'an: Xidian University Press, 1993(in Chinese)
(焦李成. 神经网络计算. 西安: 西安电子科技大学出版社, 1993)
- [20] Jiao Li-Cheng. *Application and Realization of Neural Network*. Xi'an: Xidian University Press, 1993(in Chinese)
(焦李成. 神经网络的应用与实现. 西安: 西安电子科技大学出版社, 1993)
- [21] Jiao Li-Cheng. *Neural Network System Theory*. Xi'an: Xidian University Press, 1990(in Chinese)
(焦李成. 神经网络系统理论. 西安: 西安电子科技大学出版社, 1990)
- [22] Mitchison G. The organization of sequential memory: Sparse representations and the targeting problem//*Proceedings of the Organization of Neural Networks: VCH Verlags-Gesellschaft*. Weinheim, Germany, 1988; 347-367
- [23] Rolls E, Treves A. The relative advantages of sparse versus distributed encoding for associative neuronal networks in the brain. *Network*, 1990, 1(4): 407-421
- [24] Hyvarinen A, Hoyer P O. A two layer sparse coding model learns simple and complex cell receptive fields and topography from natural images. *Vision Research*, 2001, 41(18): 2413-2423
- [25] Olshausen B A, Field D J. Sparse coding with an over complete basis set: A strategy employed by V1. *Vision Research*, 1997, 37(5): 3311-3325
- [26] Hoyer P O, Hyvarinen A. A multi-layer sparse coding network learns contour coding from natural images. *Vision Research*, 2002, 42(12): 1593-1605
- [27] Theriault C, Thome N, Cord M. HMAX-S: Deep scale representation for biologically inspired image categorization//*Proceedings of the 2011 18th IEEE International Conference on Image Processing*. Saarbrucken, Germany, 2011: 1261-1264
- [28] Baddeley R, Abbott L F, Booth M C A, et al. Responses of neurons in primary and inferior temporal visual cortices to natural scenes//*Proceedings of the Royal Society B-Biological Sciences*. London, Britain, 1997: 1775-1783
- [29] Carlson E T, Rasquinha R J, Zhang K, Connor C E. A sparse object coding scheme in area V4. *Current Biology*, 2011, 21(3): 288-293
- [30] Quian Q R, Reddy L, Kreiman G, et al. Invariant visual representation by single neurons in the human brain. *Nature*, 2005, 435(14): 1102-1107
- [31] Candes E J, Romberg J, Tao T. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *Ranaon on Nformaon Hory*, 2006, 52(2): 489-509
- [32] Chartrand R. Exact reconstruction of sparse signals via nonconvex minimization. *IEEE Signal Processing Letters*, 2007, 14(10): 707-710
- [33] Chartrand R, Staneva V. Restricted isometry properties and nonconvex compressive sensing. *Inverse Problems*, 2008, 24(3): 1-14
- [34] Candès E J, Wakin M, Boyd S. Enhancing sparsity by reweighted L1 minimization. *Journal of Fourier Analysis and Applications*, 2008, 14(1): 877-905
- [35] Elad M. *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. Germany: Springer Verlag, 2010
- [36] Xu H, Caramanis C, Mannor S. Sparse algorithms are not stable: A no-free-lunch theorem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(1): 187-193
- [37] Xu Zong-Ben, Zhang Hai, Wang Yao, Chang Xiang-Yu, Liang Yong. $L_{1/2}$ regularization. *Science China (Information Sciences)*, 2010, 53(6): 1159-1169
- [38] Mi Tie-Bin, Li Shi-Dong, Liu Yu-Long. The L1 analysis approach by sparse dual frames for sparse signal recovery represented by frames. *IEEE International Symposium on Information Theory*, 2012, 6(12): 2037-2041
- [39] Nam S, Davies M E, Elad M, Gribonval R. The cosparsity analysis model and algorithms. *Applied and Computational Harmonic Analysis*, 2013, 34(1): 30-56
- [40] Elad M, Milanfar P, Rubinstein R. Analysis versus synthesis in signal priors. *Inverse Problems*, 2007, 23(3): 947-968
- [41] Joachims T, Hofmann T, Yue Y, Yu C N J. Predicting structured objects with support vector machines. *Communications of the ACM*, 2009, 52(11): 97-104
- [42] Kulesza A, Pereira F. Structured learning with approximate inference. *Advances in Neural Information Processing Systems*, 2007, 11(1): 1-8
- [43] Liu Fang, Wu Jiao, Yang Shu-Yuan, Jiao Li-Cheng. Research advances on structured compressive sensing. *Acta Automatica Sinica*, 2013, 39(12): 1980-1995(in Chinese)
(刘芳, 武娇, 杨淑媛, 焦李成. 结构化压缩感知研究进展. *自动化学报*, 2013, 39(12): 1980-1995)
- [44] Sun Hong, Zhang Zhi-Lin, Yu Lei. From sparsity to structured sparsity: Bayesian perspective. *Signal Processing*, 2012, 6(5): 759-773(in Chinese)
(孙洪, 张智林, 余磊. 从稀疏到结构化稀疏: 贝叶斯方法. *信号处理*, 2012, 6(5): 759-773)
- [45] Mairal J, Jenatton R, Obozinski G, Bach F. Convex and network flow optimization for structured sparsity. *Journal of Machine Learning Research*, 2011, 12(9): 2681-2720
- [46] Jenatton R, Audibert J Y, Bach F. Structured variable selection with sparsity-inducing norms. *Journal of Machine Learning Research*, 2011, 12(10): 2777-2824

- [47] Huang J, Zhang T, Metaxas D. Learning with structured sparsity. *The Journal of Machine Learning Research*, 2009, 12(1): 3371-3412
- [48] Koray K, MarcAurelio R, Rob F, Yann L C. Learning invariant features through topographic filter maps//Proceedings of the International Conference on Computer Vision and Pattern Recognition. Miami, USA, 2009: 1605-1612
- [49] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 2012, 12(1): 1-9
- [50] Fawzi A, Davies M, Frossard P. Dictionary learning for fast classification based on soft-thresholding. *International Journal of Computer Vision*, 2014, 21(2): 1-16
- [51] Elad M. Sparse and redundant representation modeling—What next? *IEEE Signal Processing Letters*, 2012, 19(12): 922-928
- [52] Faul A C, Tipping M E. Analysis of sparse Bayesian learning. *Advances in Neural Information Processing Systems*, 2001, 20(3): 383-389
- [53] Allen G I. Sparse and functional principal components analysis. E Print arXiv, 2013, 9(11): 1-21
- [54] Ma Shi-Qian. Algorithms for Sparse and Low-Rank Optimization: Convergence, Complexity and Applications [Ph. D. dissertation]. Columbia University, New York, 2011
- [55] Huang Bo. Convex Optimization Algorithms and Recovery Theories for Sparse Models in Machine Learning [Ph. D. dissertation]. Columbia University, New York, 2014
- [56] Liu H, Taniguchi T. Feature extraction and pattern recognition for human motion by a deep sparse autoencoder//Proceedings of the 2014 IEEE International Conference on Computer and Information Technology. Xi'an, China, 2014: 173-181
- [57] Donoho D L. Compressed sensing. *IEEE Transactions on Information Theory*, 2006, 52(4): 1289-1239
- [58] Candès E J, Romberg J K, Tao T. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 2006, 59(8): 1207-1222
- [59] Candès E J, Romberg J K, Tao T. Robust uncertainty principles: Exact signal reconstruction from highly incomplete fourier information. *IEEE Transactions on Information Theory*, 2006, 52(8): 489-509
- [60] Jiao Li-Cheng, Yang Shu-Yuan, Liu Fang, Hou Biao. Development and prospect of compressive sensing. *Acta Electronica Sinica*, 2011, 39(7): 1651-1662(in Chinese)
(焦李成, 杨淑媛, 刘芳, 侯彪. 压缩感知回顾与展望. *电子学报*, 2011, 39(7): 1651-1662)
- [61] Baraniuk R. Compressive sensing. *IEEE Signal Processing Magazine*, 2007, 32(7): 118-120
- [62] Lustig M, Donoho D L, Santos J M, et al. Compressed Sensing MRI. *IEEE Signal Processing Magazine*, 2008, 25(2): 72-82
- [63] Fauvel S, Ward R K. An energy efficient compressed sensing framework for the compression of electroencephalogram signals. *Sensors (Basel, Switzerland)*, 2014, 14(1): 1474-1496
- [64] Figueiredo M A T, Nowak R D, Wright S J. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE Journal of Selected Topics in Signal Processing*, 2007, 1(4): 586-597
- [65] Starck J L, Elad M, Donoho D L. Redundant multiscale transforms and their application for morphological component analysis. *Advances in Imaging and Electron Physics*, 2004, 132(12): 287-348
- [66] Farsiu S, Robinson D, Elad M, Milanfa P. Advanced and challenges in super-resolution. *The International Journal of Imaging Systems and Technology*, 2004, 14(2): 47-57
- [67] Starck J L, Elad M, Donoho D L. Image decomposition via the combination of sparse representations and a variational approach. *IEEE Transactions on Image Processing*, 2005, 14(10): 1570-1582
- [68] Donoho D L, Elad M, Temlyakov V. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Transactions on Information Theory*, 2006, 52(1): 6-18
- [69] Elad M, Starck J L, Querre P, Donoho D L. Simultaneous cartoon and texture image inpainting using morphological component analysis. *Journal on Applied and Computational Harmonic Analysis*, 2005, 19(32): 340-358
- [70] Bobin J, Moudden Y, Starck J L, Elad M. Morphological diversity and source separation. *IEEE Signal Processing Letters*, 2006, 13(7): 409-412
- [71] Elad M, Goldenberg R, Kimmel R. Low bit-rate compression of facial images. *IEEE Transactions on Image Processing*, 2007, 16(9): 2379-2383
- [72] Fadili J M, Starck J L, Elad M, Donoho D L. Reproducible research in signal and image decomposition and inpainting. *Computing in Science & Engineering*, 2010, 12(1): 44-62
- [73] Adler A, Elad M, Yacov H. Sparse coding with anomaly detection. *Journal of Signal Processing Systems*, 2015, 79(2): 179-188
- [74] Wright J, Yang A Y, Ganesh A, et al. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, 31(2): 210-227
- [75] Gao S, Tsang I W H, Chia L T. Kernel sparse representation for image classification and face recognition//Proceedings of the 11th European Conference on Computer Vision-ECCV. Crete, Greece, 2010: 1-14
- [76] Zhang L, Yang M, Feng X C. Sparse representation or collaborative representation: Which helps face recognition//Proceedings of the IEEE International Conference on Computer Vision. Barcelona, Spain, 2011: 471-478

- [77] Yang M, Zhang L, Shiu S C K, Zhang D. Robust kernel representation with statistical local features for face recognition. *IEEE Transactions on Neural Networks and Learning Systems*, 2013, 22(2): 900-912
- [78] Han Y H, Wu F, Tian Q, Zhuang Y T. Image annotation by input-output structural grouping sparsity. *IEEE Transactions on Image Processing*, 2012, 21(6): 3066-3079
- [79] Pontil M, Mourao-Miranda J, Baldassarre L. Structured sparsity models for brain decoding from fMRI data// *Proceedings of the IEEE Computer Society International Workshop on Pattern Recognition in Neuroimaging*. London, UK, 2012: 5-8
- [80] Jia K, Chan T H, Ma Y. Robust and practical face recognition via structured sparsity//Fitzgibbon A, Lazebnik S, Perona P, et al eds. *Computer Vision-ECCV*. Berlin: Springer, 2012: 331-334
- [81] Zhang H, Zhang Y, Huang T S. Exploiting Structured Sparsity for Image Deblurring//*Proceedings of the IEEE International Conference on Multimedia and Expo*. Melbourne, Australia, 2012: 616-621
- [82] Asaei A, Golbabaee M, Bourlard H, Cevher V. Structured sparsity models for reverberant speech separation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2014, 22(3): 620-633
- [83] Roos M J, Wolmetz M, Chevillet M A. A hierarchical model of vision can also recognize speech. *BMC Neuroscience*, 2014, 15(1): 187
- [84] Gu W, Xiang C, Lin H. Modified HMAX models for facial expression recognition//*Proceedings of the IEEE International Conference on Control and Automation*. Christchurch, New Zealand, 2009: 1509-1514
- [85] Zhao Hong-Wei, Cui Hong-Rui, Dai Jin-Bo, et al. Contour detection based on HMAX model and non-classical receptive field inhibition. *Journal of Jilin University*, 2012, 42(1): 128-133(in Chinese)
(赵宏伟, 崔弘睿, 戴金波等. 基于HMAX模型和非经典感受野抑制的轮廓提取. *吉林大学学报*, 2012, 42(1): 128-133)
- [86] Hong Hanh P T, Ngoc L Q. Multiple objects detection on street using Hmax features and color clue//*Proceedings of the 2012 IEEE International Symposium on Signal Processing and Information Technology*. Ho Chi Minh, Vietnam, 2012: 90-94
- [87] Yaghoubi Z, Faez K, Eliasi M. Face recognition using HMAX method for feature extraction and support vector machine classifier//*Proceedings of the 24th International Conference Image and Vision Computing*. Wellington, New Zealand, 2009: 421-424
- [88] Lazebnik S, Schmid C, Ponce J. Beyond bags of features: Spatial pyramid matching for recognizing nature scene categories//*Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. New York, USA, 2006: 2169-2178
- [89] Roweis S, Ghahramani Z. A unifying review of linear Gaussian models. *Neural Computation*, 1999, 11(2): 305-345
- [90] Bengio Y. Learning deep architectures for AI. *Foundations & Trends in Machine Learning*, 2009, 2(1): 1-127
- [91] Eckhorn R, Bauer R, Jordan W, et al. Coherent oscillations: A mechanism of feature linking in the visual cortex, multiple electrode and correlation analyses in the cat. *Biological Cybernetics*, 1988, 60(2): 121-130
- [92] Grill-Spector K, Malach R. The human visual cortex. *Annual Review of Neuroscience*, 2004, 27(1): 649-677
- [93] Kivinen J, Smola A J, Williamson R C. Learning with Kernels. *IEEE Transactions on Signal Processing*, 2002, 52(8): 2165-2176
- [94] Arnold L, Olivier Y. Layer-wise learning of deep generative models. *HAL-INRIA*, 2012, 14(3): 1-43
- [95] Rubanov N S. The layer-wise method and the backpropagation hybrid approach to learning a feedforward neural network. *IEEE Transactions on Neural Network*, 2000, 11(2): 295-305
- [96] Jiao Li-Cheng, Bao Zheng. Advances and prospective in wavelet theory and its application. *Acta Electronica Sinica*, 1993, 21(7): 91-96(in Chinese)
(焦李成, 保铮. 子波理论与应用: 进展与展望. *电子学报*, 1993, 21(7): 91-96)
- [97] Jiao Li-Cheng, Sun Qiang. Advances and prospective on image perception and recognition in multiscale transform domains. *Chinese Journal of Computers*, 2006, 29(2): 177-193(in Chinese)
(焦李成, 孙强. 多尺度变换域图像的感知与识别: 进展和展望. *计算机学报*, 2006, 29(2): 177-193)
- [98] Jiao Li-Cheng, Tan Shan. Development and prospect of image multiscale geometric analysis. *Acta Electronica Sinica*, 2003, 31(12): 1975-1981(in Chinese)
(焦李成, 谭山. 图像的多尺度几何分析: 回顾和展望. *电子学报*, 2003, 31(12): 1975-1981)
- [99] Jiao Li-Cheng, Yang Shu-Yuan. *Adaptive Multi-Scale Network Theory and Application*. Beijing: Science Press, 2008 (in Chinese)
(焦李成, 杨淑媛. 自适应多尺度网络理论与应用. 北京: 科学出版社, 2008)
- [100] Huang Kai-Qi, Tan Tie-Niu. Review on computational model for vision. *Pattern Recognition and Artificial Intelligence*, 2013, 26(10): 951-958(in Chinese)
(黄凯奇, 谭铁牛. 视觉认知计算模型综述. 模式识别与人工智能, 2013, 26(10): 951-958)
- [101] Huang Y Z, Huang K Q, Tao D C, et al. Enhanced biologically inspired model for object recognition. *IEEE Transactions on Systems, Man and Cybernetics*, 2011, 41(6): 1668-1680



JIAO Li-Cheng, born in 1959, Ph. D. , professor. His research interests include intelligent perception and image understanding.

ZHAO Jin, born in 1986, Ph. D. candidate. His research interests include sparse representation and

modeling.

YANG Shu-Yuan, born in 1978, Ph. D. , professor. Her research interests include intelligent signal and image processing, machine learning.

LIU Fang, born in 1963, professor. Her research interests include intelligent signal processing, pattern recognition.

XIE Wen, born in 1989, Ph. D. candidate. Her research interests include PoL-SAR image classification.

Background

Under the explosive increase of global data in various fields, the term of big data is mainly used to describe massive data sets having large, more varied and complex structure with difficulties of storing, analyzing and visualizing for further process or results. Compared with traditional datasets, big data typically means masses of unstructured data that need more real time analysis. How to effectively organize and manage such datasets have become a very important problem. Many theories, models and methods have been proposed and extensively studied. Sparse Cognitive Learning, Computing and Recognition, which is based on the understanding of biological sparse cognitive mechanism to establish a computation model, aims to simulate intelligent biological visual system and complete the recognition task under complex scene. It has been applied successfully in many fields, such

as pattern for dealing with big data.

This paper provides a review of the main research advances and applications of Sparse Cognitive Learning, Computing and Recognition. The authors summarize three different forms of this paradigm, such as sparse coding, structured sparse and hierarchy sparse, and describe the relationships between three different forms in detail. Finally, point out some problems of model and research directions.

This work is partially supported by the National Basic Research Program of China (No.2013CB329402), the Programme of Introducing Talents of Discipline to Universities (No. B07048), the Major National Scientific Research Projects (Nos. 91438201 and 91438103), the Program for Changjiang Scholars and Innovative Research Team in University (No. IRT1170).